GDC Data Portal User's Guide

NCI Genomic Data Commons (GDC)

Contents

1	Getting Started	9
	Getting Started	9
	The GDC Data Portal: An Overview	9
	Accessing the GDC Data Portal	9
	Views	10
	Toolbar	11
	Tables	12
	Table Sort	12
	Table Arrangement	13
	Table Size	13
	Table Export	13
	Filtering and Searching	13
	Facet Filters	13
	Quick Search	15
	Advanced Search	18
	Manage Sets	18
	Upload Sets	18
	Export Sets	19
	Review Sets	20
	Creating Sets from GDC Portal Filters	20
2	Projects	21
	Projects	21
	Projects Page	21
	Visualizations	22
	Top Mutated Cancer Genes in Selected Projects	22
	Case Distribution per Project	23
	Projects Table	23
	Projects Graph	23
	Facets Panel	24
	Project Summary Page	25

3	Exploration	27
	Exploration	27
	Filters / Facets	27
	Case Filters	28
	Clinical Filters	30
	Gene Filters	32
	Mutation Filters	34
	Results	36
	Cases	36
	Case Summary Page	37
	Clinical and Biospecimen Information	38
	Biospecimen Search	42
	Most Frequent Somatic Mutations for a Case	43
	Genes	43
	Gene Summary Page	44
	External References	45
	Cancer Distribution	45
	Protein Viewer	46
	Most Frequent Somatic Mutations	47
	Mutations	47
	Mutation Summary Page	48
	External References	49
	Consequences	49
	Cancer Distribution	49
	Protein Viewer	50
	OncoGrid	51
	OncoGrid Options	52
	OncoGrid Color Picker	52
	File Navigation	53
	Survival Analysis	54
	D. V	-
Ł	Repository	56
	Repository	56
	Filters / Facets	56
	Facets Panel	57
	Adding Custom Facets	59
	Annotations View	60
	Results	62
	Navigation	62

	Files List	63
	File Summary Page	65
	BAM Slicing	67
	Cases List	67
	Image Viewer	68
	How to Access the Image Viewer	69
	Image Viewer Features	71
5	Genes and Mutations	73
	Gene and Mutation Summary Pages	73
	Gene Summary Page	73
	Summary	73
	External References	74
	Cancer Distribution	74
	Protein Viewer	75
	Most Frequent Mutations	75
	Mutation Summary Page	76
	Summary	76
	External References	76
	Consequences	76
	Cancer Distribution	77
	Protein Viewer	78
6	Custom Set Analysis	79
	Analysis	79
	Generating a Cohort for Analysis	79
	Upload Case Set	80
	Upload Gene Set	82
	Upload Mutation Set	83
	Analysis Page	85
	Analysis Page: Set Operations	85
	Analysis Tab: Cohort Comparison	86
	Analysis Tab: Clinical Data Analysis	87
	Selecting a Case Set	87
	Enabling Clinical Variable Cards	88
	Exploring Clinical Card Visualizations	92
	Histogram	92
	Survival Plot	93
	Box Plot & QQ Plot	94
	Creating Custom Bins	95

	Categorical Binning	96
	Continuous Binning	97
	Other Useful Functions	99
	Analysis Page: Results	100
7	Annotations	101
	Annotations	101
	Annotations View	101
	Facets Panel	102
	Annotation Categories and Classification	102
	Annotation Detail Page	102
8	Advanced Search	104
	Advanced Search	104
	Overview: GQL	104
	Switching between Advanced Search and Facet Filters	105
	Using the Advanced Search	106
	Auto-complete	106
	Field Auto-complete	106
	Value Auto-complete	106
	Setting Precedence of Operators	107
	Keywords	107
	" AND " Keyword	107
	"OR" Keyword	108
	Operators	108
	List of Operators and Query format	108
	"=" Operator - EQUAL	108
	"!=" Operator - NOT EQUAL	109
	">" Operator - GREATER THAN	109
	">=" Operator - GREATER THAN OR EQUALS	109
	"<" Operator - LESS THAN	109
	"<=" Operator - LESS THAN OR EQUALS	110
	"IN" Operator	110
	"EXCLUDE" Operator	110
	"IS MISSING" Operator	111
	"NOT MISSING" Operator	111
	Special Cases	111
	Date Format	111
	Using Quotes	111
	Age at Diagnosis - Unit in Days	111
	Fields Reference	111

9	Authentication	112
	Authentication	112
	Overview	112
	Logging into the GDC	112
	User Profile	114
	GDC Authentication Tokens	114
	Logging Out	115
10	File Cart	116
	Cart and File Download	116
	GDC Cart	116
	Cart Summary	116
	Cart Items	117
	Download Options	117
	GDC Data Transfer Tool	118
	Controlled Files	118
	Authentication	119
	Logging into the GDC	119
	User Profile	121
	GDC Authentication Tokens	121
	Logging Out	122
11	. Image Viewer	123
	Image Viewer	123
	How to access the image viewer	123
	Image viewer features	126
	Example of navigation to the image viewer	127
12	Legacy Archive	129
	Legacy Archive	129
	Overview	129
	File Page	130
	Archive	131
	Metadata Files	131
	File Cart	131

3 Release Notes	132
Data Portal Release Notes	132
Release 1.25.1	133
New Features and Changes	133
Bugs Fixed Since Last Release	133
Known Issues and Workarounds	133
Release 1.25.0	134
New Features and Changes	134
Bugs Fixed Since Last Release	134
Known Issues and Workarounds	134
Release 1.24.1	135
New Features and Changes	135
Bugs Fixed Since Last Release	135
Known Issues and Workarounds	136
Release 1.23.1	137
New Features and Changes	137
Bugs Fixed Since Last Release	137
Known Issues and Workarounds	137
Release 1.23.0	137
New Features and Changes	138
Bugs Fixed Since Last Release	138
Known Issues and Workarounds	138
Release 1.22.0	139
New Features and Changes	139
Bugs Fixed Since Last Release	139
Known Issues and Workarounds	139
Release 1.21.0	140
New Features and Changes	140
Bugs Fixed Since Last Release	140
Known Issues and Workarounds	140
Release 1.20.0	141
New Features and Changes	141
Bugs Fixed Since Last Release	141
Known Issues and Workarounds	141
Release 1.19.0	142
New Features and Changes	142
Bugs Fixed Since Last Release	142
Known Issues and Workarounds	142
Release 1.18.0	143

New Features and Changes	43
Bugs Fixed Since Last Release	43
Known Issues and Workarounds	43
Release 1.17.0	44
New Features and Changes	44
Bugs Fixed Since Last Release	44
Known Issues and Workarounds	44
Release 1.16.0	44
New Features and Changes	44
Bugs Fixed Since Last Release	45
Known Issues and Workarounds	45
Release 1.15.0	45
New Features and Changes	45
Bugs Fixed Since Last Release	46
Known Issues and Workarounds	46
Release 1.14.0	46
New Features and Changes	46
Bugs Fixed Since Last Release	47
Known Issues and Workarounds	47
Release 1.13.0	47
New Features and Changes	47
Bugs Fixed Since Last Release	47
Known Issues and Workarounds	48
Release 1.12.0	48
New Features and Changes	48
Bugs Fixed Since Last Release	48
Known Issues and Workarounds	49
Release 1.11.0	49
New Features and Changes	50
Bugs Fixed Since Last Release	50
Known Issues and Workarounds	50
Release 1.10.0	51
New Features and Changes	51
Bugs Fixed Since Last Release	51
Known Issues and Workarounds	51
Release 1.9.0	52
New Features and Changes	52
Bugs Fixed Since Last Release	52
Known Issues and Workarounds	52

Release 1.8.0	152
New Features and Changes	153
Bugs Fixed Since Last Release	153
Known Issues and Workarounds	153
Release 1.6.0	154
New Features and Changes	154
Bugs Fixed Since Last Release	154
Known Issues and Workarounds	155
Release 1.5.2	155
New Features and Changes	155
Bugs Fixed Since Last Release	156
Known Issues and Workarounds	156
Release 1.4.1	156
New Features and Changes	156
Bugs Fixed Since Last Release	157
Known Issues and Workarounds	157
Release 1.3.0	157
New Features and Changes	158
Bugs Fixed Since Last Release	158
Known Issues and Workarounds	158
Release 1.2.0	159
New Features and Changes	159
Bugs Fixed Since Last Release	159
Release 1.1.0	160
New Features and Changes	160
Bugs Fixed Since Last Release	160
Known Issues and Workarounds	160
Release 1.0.1	161
New Features and Changes	161
Bugs Fixed Since Last Release	161
Known Issues and Workarounds	162

Chapter 1

Getting Started

Getting Started

The GDC Data Portal: An Overview

The Genomic Data Commons (GDC) Data Portal provides users with web-based access to data from cancer genomics studies. Key GDC Data Portal features include:

- Open, granular access to information about all datasets available in the GDC.
- Advanced search and visualization-assisted filtering of data files.
- Data visualization tools to support the analysis and exploration of data (including on a gene and mutation level from Open-Access MAF files).
- Cart for collecting data files of interest.
- Authentication using eRA Commons credentials and auathorization using dbGaP for access to controlled data files.
- Secure data download directly from the cart or using the GDC Data Transfer Tool.

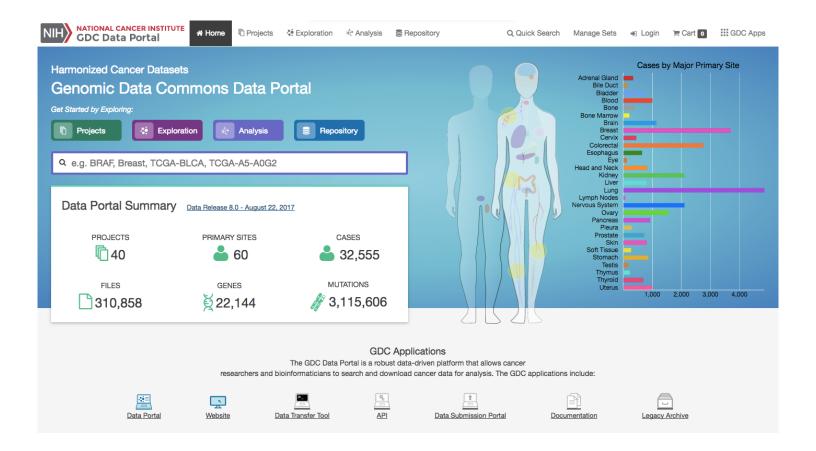
For more information about available datasets, see the GDC Website.

Accessing the GDC Data Portal

The GDC Data Portal is accessible using a web browser such as Chrome, Firefox, and Microsoft Edge at the following URL:

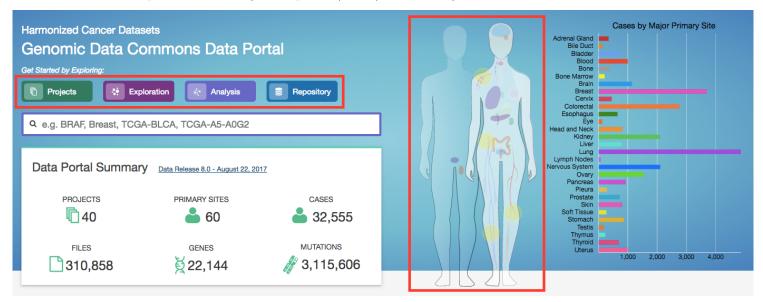
https://portal.gdc.cancer.gov

The front page displays a summary of all available datasets:



Views

The GDC Data Portal provides five navigation options (Views) for browsing available harmonized datasets:



- **Projects**: The Projects link directs users to the Projects Page, which gives an overall summary of project-level information, including the available data for each project.
- Exploration: The Exploration link takes users to the Exploration Page, which allows users to explore data by utilizing various case, genes and mutation filters.

- Analysis: The Analysis link directs users to the Analysis Page. This page has features available for users to compare different cohorts or analyze the clinical variables of a specific cohort. These cohorts can either be generated with existing filters (e.g. males with lung cancer) or through custom selection.
- Repository: The Repository link directs users to the Repository Page. Here users can see the data files available for download at the GDC and apply file/case filters to narrow down their search.
- Human Outline: The home page displays a human anatomical outline that can be used to refine their search. Choosing an associated organ will direct the user to a listing of all projects associated with that primary site. For example, clicking on the human brain will show only cases and projects associated with brain cancer (TCGA-GBM and TCGA-LGG). The number of cases associated with each primary site is also displayed here and separated by project.

Each view provides a distinct representation of the same underlying set of GDC data and metadata. The GDC also provides access to certain unharmonized data files generated by GDC-hosted projects. These files and their associated metadata are not represented in the views above; instead they can be found in the GDC Legacy Archive.

The Projects, Exploration, Analysis and Repository pages can be accessed from the GDC Data Portal front page and from the toolbar (see below). The annotations view is accessible from Repository view. A link to the GDC Legacy Archive is available on the GDC Data Portal front page and in the GDC Apps menu (see below).

Toolbar

The toolbar available at the top of all pages in the GDC Data Portal provides convenient navigation links and access to authentication and quick search.

The left portion of this toolbar provides access to the Home Page, Projects Page, Exploration Page, Analysis Page, and a link to Repository Page:









Analysis 🖈

Repository

The right portion of this toolbar provides access to quick search, and the GDC Apps menu:

Q Quick Search

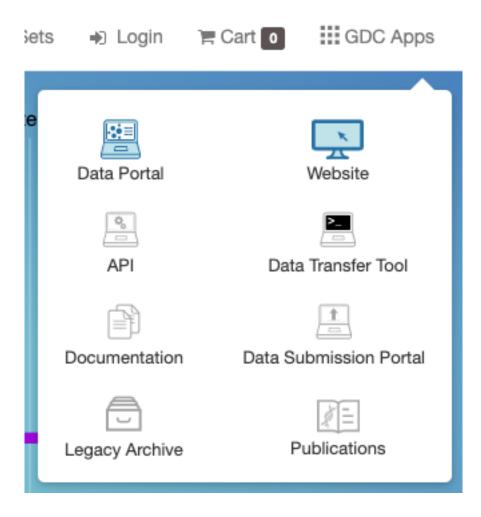
Manage Sets

→ Login

🗏 Cart 0

GDC Apps

The GDC Apps menu provides links to all resources provided by the GDC, including the GDC Legacy Archive.



Tables

Tabular listings are the primary method of representing available data in the GDC Data Portal. Tables are available in all views and in the file cart. Users can customize each table by specifying columns, size, and sorting.

Table Sort

The sort button is available in the top right corner of each table. To sort by a column, place a checkmark next to it and select the preferred sort direction. If multiple columns are selected for sorting, data is sorted column-by-column in the order that the columns appear in the sort menu: the topmost selected column becomes the primary sorting parameter; the selected column below it is used for secondary sort, etc.

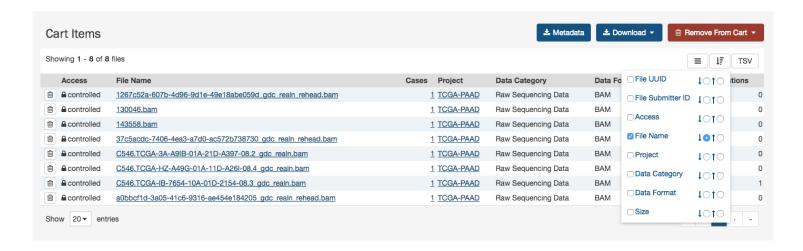


Table Arrangement

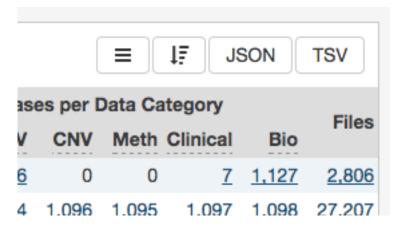
The arrange button allows users to adjust the order of columns in the table and select which columns are displayed.

Table Size

Table size can be adjusted using the menu in the bottom left corner of the table. The menu sets the maximum number of rows to display. If the number of entries to be displayed exceeds the maximum number of rows, then the table will be paginated, and navigation buttons will be provided in the bottom right corner of the table to navigate between pages.

Table Export

In the Repository, Projects, and Annotations views, tables can be exported in either a JSON or TSV format. The JSON button will export the entire table's contents into a JSON file. The TSV button will export the current view of the table into a TSV file.



Filtering and Searching

The GDC Data Portal offers three different means of searching and filtering the available data: facet filters, quick search, and advanced search.

Facet Filters

Facets on the left of each view (Projects, Exploration, and Repository) represent properties of the data that can be used for filtering. Some of the available facets are project name, disease type, patient gender and age at diagnosis, and various data

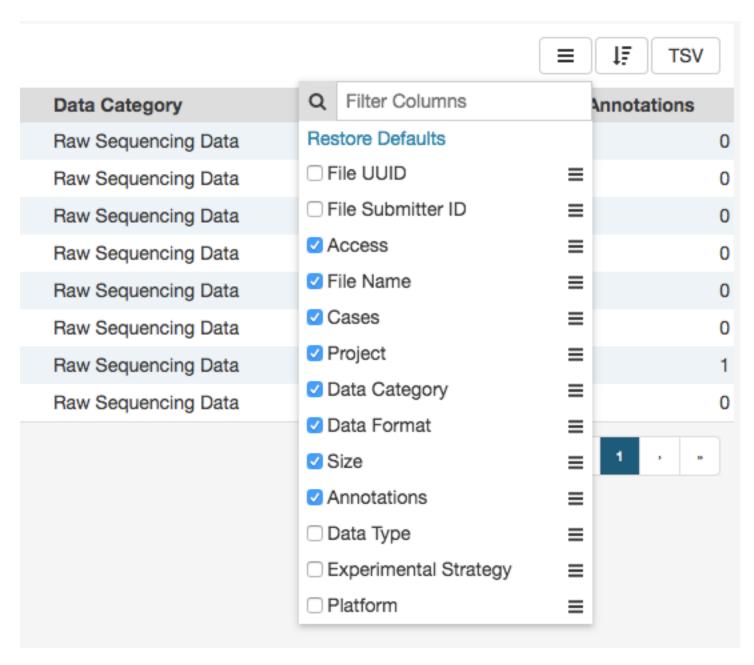


Figure 1.1: Selecting table columns



Figure 1.2: Specifying table size

formats and categories. Each facet displays the name of the data property, the available values, and numbers of matching entities for each value (files, cases, mutations, genes, annotations, or projects, depending on the context).

Below are two file facets available in the Repository view. A Data Type facet filter is applied, filtering for "Aligned Reads" files.

Multiple selections within a facet are treated as an "OR" query: e.g. "Aligned Reads" OR "Annotated Somatic Mutation". Selections in different facets are treated as "AND" queries: e.g. Data Type: "Aligned Reads" AND Experimental Strategy: "RNA-Seq".

The information displayed in each facet reflects this: in the example above, marking the "Aligned Reads" checkbox does not change the numbers or the available values in the *Data Type* facet where the checkbox is found, but it does change the values available in the *Experimental Strategy* facet. The *Experimental Strategy* facet now displays only values from files of *Data Type* "Aligned Reads".

Custom facet filters can be added in the Repository View to expand the GDC Data Portal's filtering capabilities.

Quick Search

The quick search feature allows users to find cases, files, mutations, or genes using a search query (i.e. UUID, filename, gene name, DNA Change, project name, id, disease type or primary site). Quick search is available by clicking on the magnifier in the right section of the toolbar (which appears on every page) or by using the search bar on the Home Page.

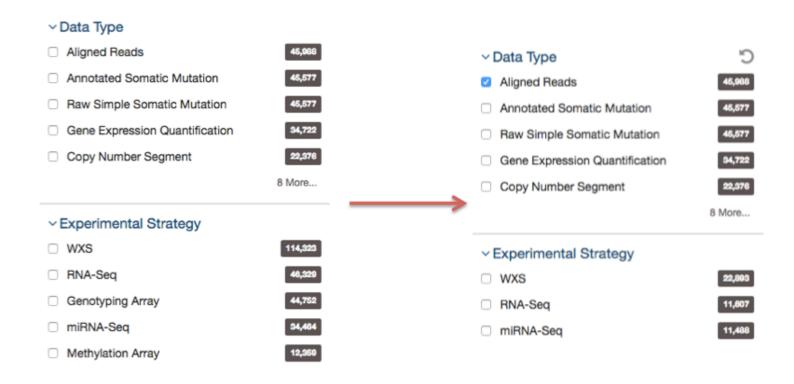
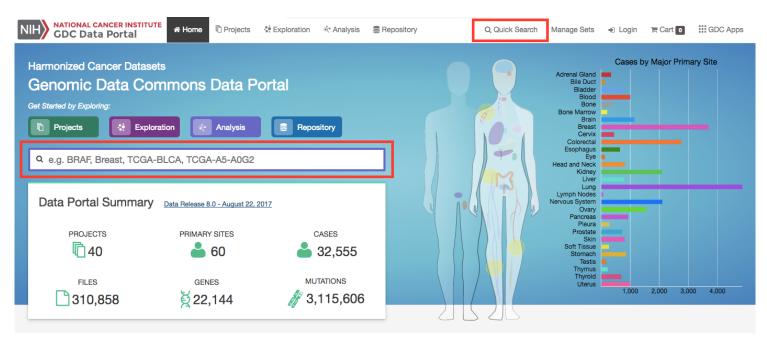
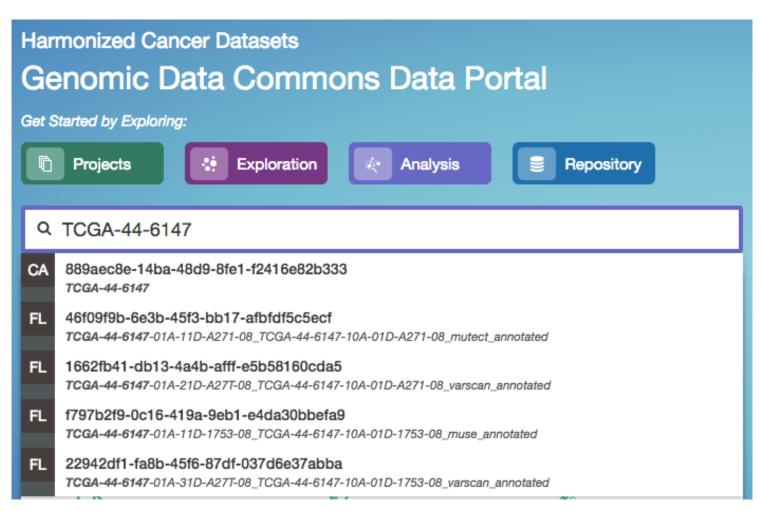


Figure 1.3: Facets (no filter applied)



Search results are displayed as the user is typing, with labels indicating the type of each search result in the list (project, case, or file). Users will see a brief description of the search results, which may include the UUID, submitter ID, or file name. Clicking on a selected result or pressing enter will open a detail page with additional information.

Home Page Quick Search:



Toolbar Quick Search:





Advanced Search

Advanced Search is available in Repository View. It allows users to construct complex queries with a custom query language and auto-complete suggestions. See Advanced Search for details.

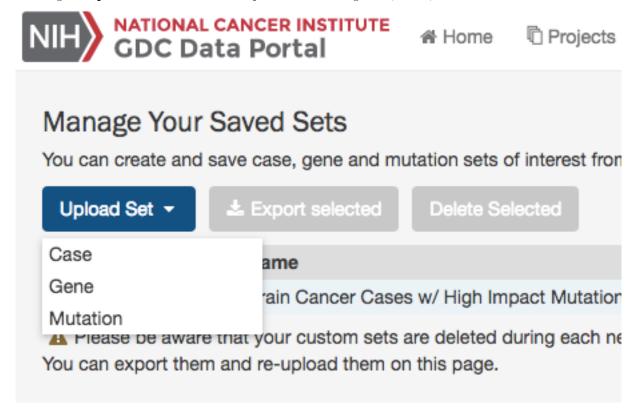
Manage Sets

The Manage Sets button at the top of the GDC Portal stores sets of cases, genes, or mutations of interest. On this page, users can review the sets that have been saved as well as upload new sets and delete existing sets.

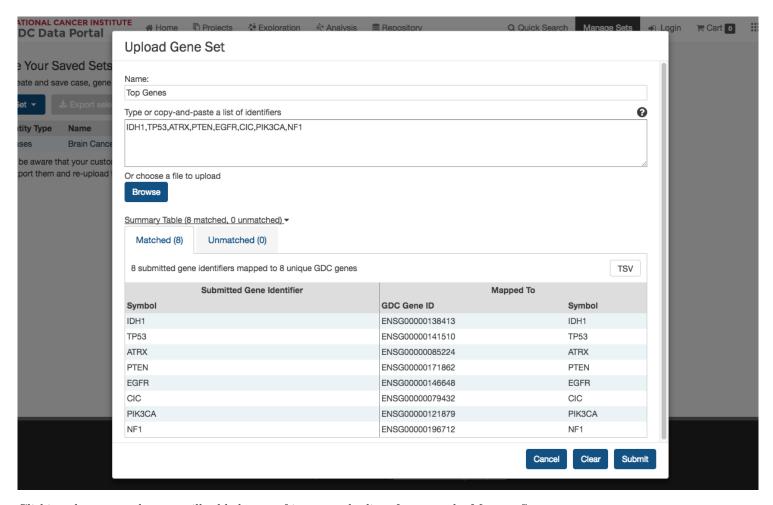


Upload Sets

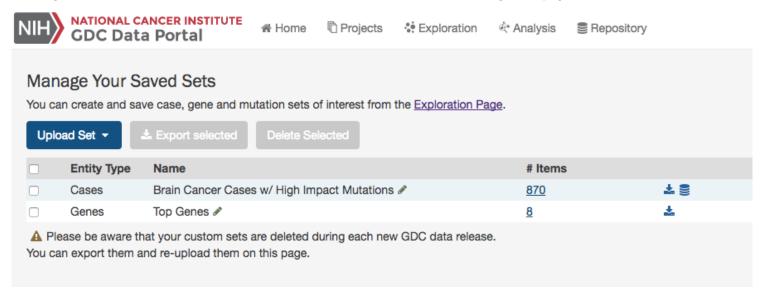
Clicking the Upload Set button shows options for creating Case, Gene, or Mutation sets.



Upon clicking one of the menu items, users are shown a dialog where they can enter unique identifiers (i.e. UUIDs, TCGA Barcodes, gene symbols, mutation UUIDs, etc.) that describe the set.

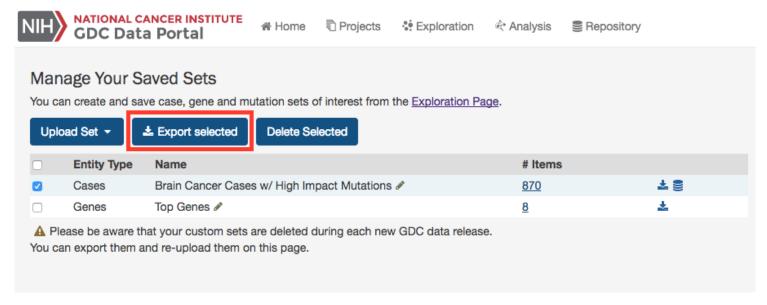


Clicking the Submit button will add the set of items to the list of sets on the Manage Sets page.



Export Sets

Users can export selected sets on this page by first clicking the checkboxes next to each set, then clicking the Export selected button at the top of the table.



A text file containing the UUID of each case, gene or mutation is downloaded after clicking this button.

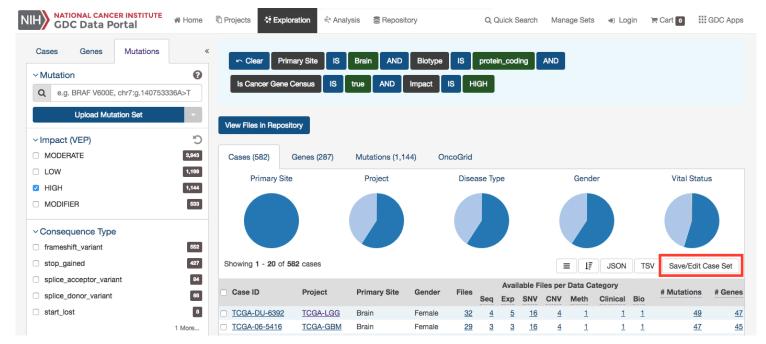
Review Sets

There are a few buttons in the list of sets that allows a user to get further information about each one.

- ___ Items___: Clicking the link under the Items column navigates the user to the Exploration page using the set as a filter.
- **Download/View**: To the right of the Items column are buttons that will download the list as a TSV or open the cases in the Repository Page.

Creating Sets from GDC Portal Filters

Many pages on the GDC Portal have an option called Save Sets that allows users to save a group of cases, mutations, or genes for further analysis. After using the filtering options on the Exploration Page as an example, users can click the Save Case/Gene/Mutation Set button to save this set.



Chapter 2

Projects

Projects

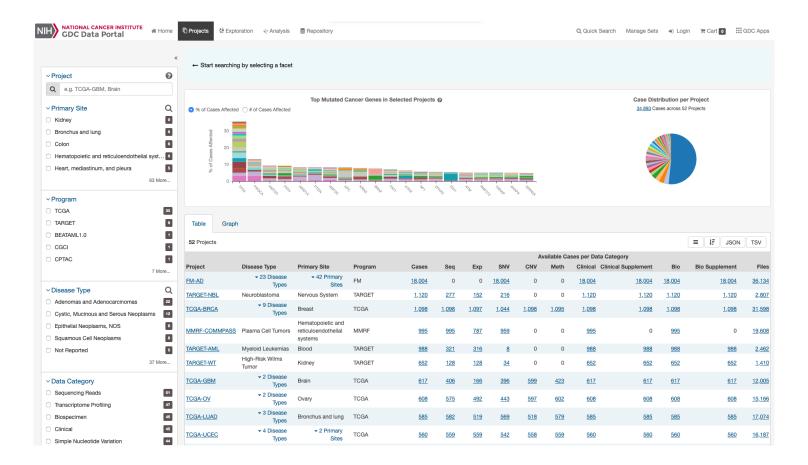
At a high level, data in the Genomic Data Commons is organized by project. Typically, a project is a specific effort to look at particular type(s) of cancer undertaken as part of a larger cancer research program. The GDC Data Portal allows users to access aggregate project-level information via the Projects Page and Project Summary Pages.

Projects Page

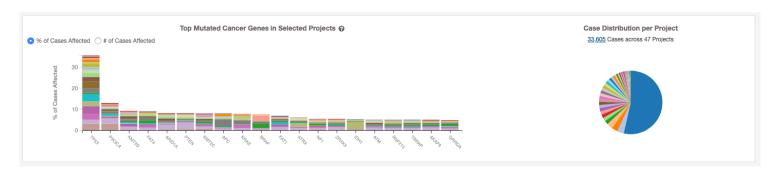
The Projects Page provides an overview of all harmonized data available in the Genomic Data Commons, organized by project. It also provides filtering, navigation, and advanced visualization features that allow users to identify and browse projects of interest. Users can access the Projects Page from the GDC Data Portal Home page or from the Data Portal toolbar.

On the left, a panel of facets allow users to apply filters to find projects of interest. When facet filters are applied, the table and visualizations on the right are updated to display only the matching projects. When no filters are applied, all projects are displayed.

The right side of the Projects Page displays a few visualizations of the data (Top Mutated Genes in Selected Projects and Case Distribution per Project). Below these graphs is a table that contains a list of projects and select details about each project, such as the number of cases and data files. The Graph tab provides a visual representation of this information.



Visualizations



Top Mutated Cancer Genes in Selected Projects

This dynamically generated bar graph shows the 20 genes with the most mutations across all projects. The genes are filtered by those that are part of the Cancer Gene Census and that have the following types of mutations: missense_variant, frameshift_variant, start_lost, stop_lost, initiator_codon_variant, and stop_gained. The bars represent the frequency of mutations per gene and is broken down into different colored segments by project. The graphic is updated as filters are applied for projects, programs, disease types, and data categories available in the project.

Note: Due to these filters, the number of cases displayed here will be less that the total number of cases per project.

Hovering the cursor over each bar will display information about the number of cases affected by the disease type and clicking on each bar will launch the Gene Summary Page for the gene associated with the mutation.

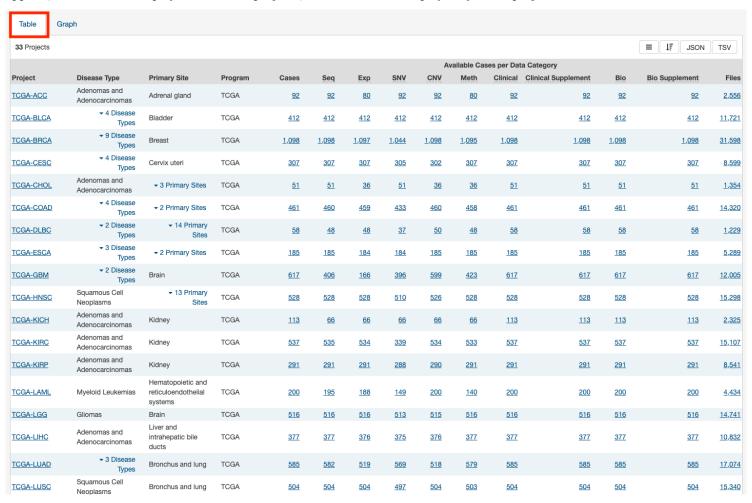
Users can toggle the Y-Axis of this bar graph between a percentage or raw number of cases affected.

Case Distribution per Project

A pie chart displays the relative number of cases for each project. Hovering the cursor over each portion of the graph will display the project with the number of associated cases. Filtering projects at the left panel will update the pie chart.

Projects Table

The Table tab lists projects by Project ID and provides additional information about each project. If no facet filters have been applied, the table will display all available projects; otherwise it will display only those projects that match the selected criteria.

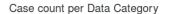


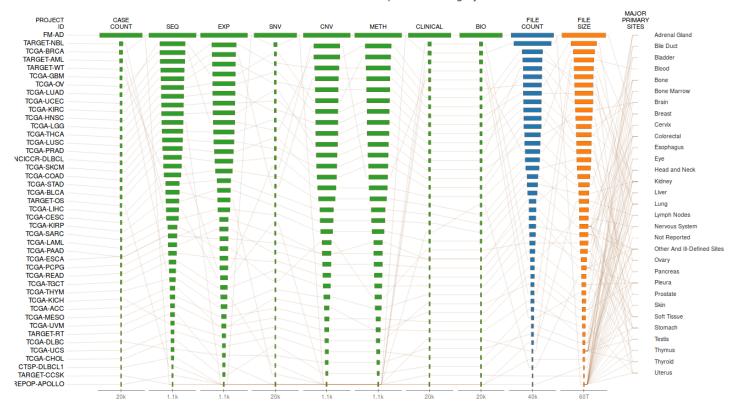
The table provides links to Project Summary Pages.

Projects Graph

The Graph tab contains an interactive view of information in the Table tab. The numerical values in Case Count, File Count, and File Size columns are represented by bars of varying length according to size. These columns are sorted independently in descending order. Mousing over an element of the graph connects it to associated elements in other columns, including Project ID and major Primary Sites.



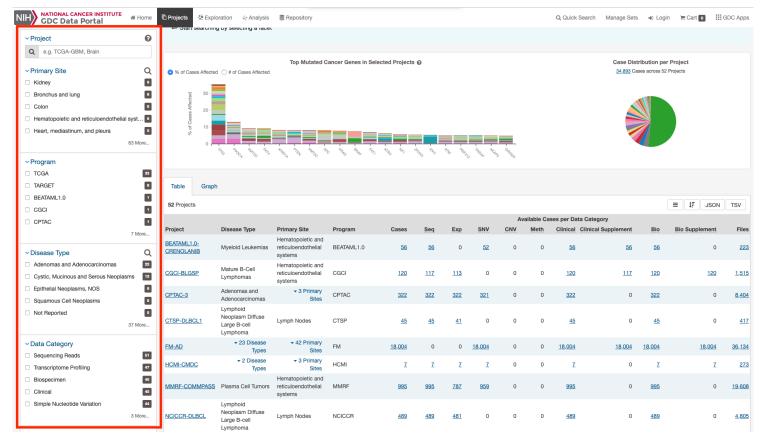




Most elements in the graph are clickable, allowing the user to open the associated cases or files in Repository Page. Like the projects table, the graph will reflect any applied facet filters.

Facets Panel

Facets represent properties of the data that can be used for filtering. The facets panel on the left allows users to filter the projects presented in the Table and Graph tabs as well as visualizations.



Users can filter by the following facets:

- Project: Individual project ID.
- Primary Site: Anatomical site of the cancer under investigation or review.
- **Program**: Research program that the project is part of.
- Disease Type: Type of cancer studied.
- Data Category: Type of data available in the project.
- Experimental Strategy: Experimental strategies used for molecular characterization of the cancer.

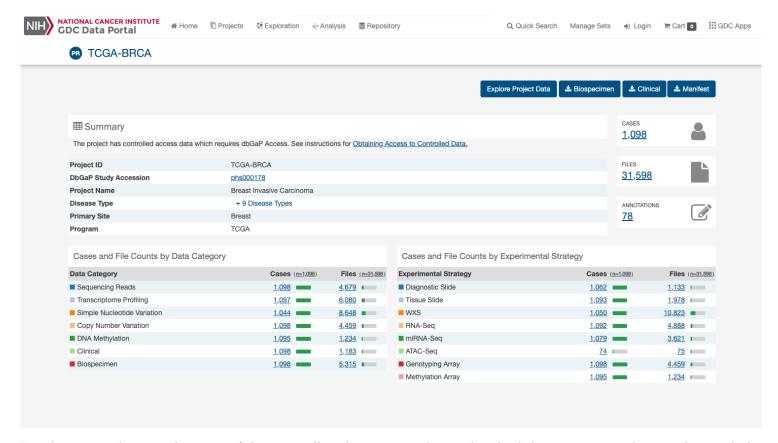
Filters can be applied by selecting values of interest in the available facets, for example "WXS" and "RNA-Seq" in the "Experimental Strategy" facet and "Brain" in the "Primary Site" facet. When facet filters are applied, the Table and Graph tabs are updated to display matching projects, and the banner above the tabs summarizes the applied filters. The banner allows the user to click on filter elements to remove the associated filters and includes a link to view the matching cases and files.



For information on how to use facet filters, see Getting Started.

Project Summary Page

Each project has a Summary Page that provides an overview of all available cases, files, and annotations available. Clicking on the numbers in the summary table will display the corresponding data.



Four buttons in the top right corner of the screen allow the user to explore or download the entire project dataset, along with the associated project metadata:

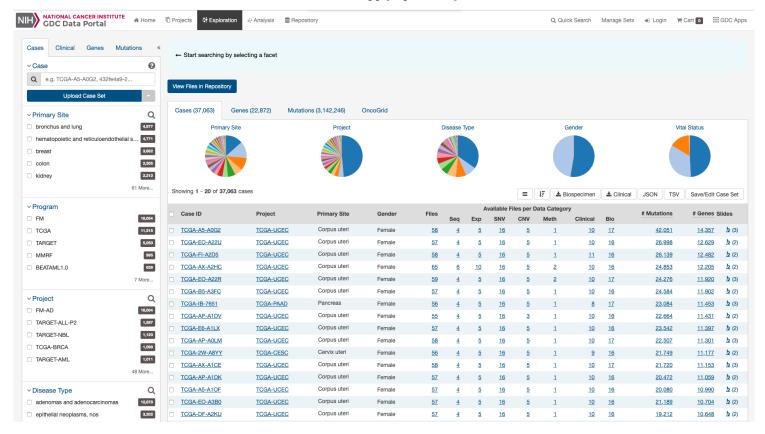
- Explore Project Data: Opens Exploration Page with summary project information.
- Biospecimen: Downloads biospecimen metadata associated with all cases in the project in either TSV or JSON format.
- Clinical: Downloads clinical metadata about all cases in the project in either TSV or JSON format.
- Manifest: Downloads a manifest for all data files available in the project. The manifest can be used with the GDC Data Transfer Tool to download the files.

Chapter 3

Exploration

Exploration

The Exploration Page allows users to explore data in the GDC using advanced filters/facets, which includes those on a gene and mutation level. Users choose filters on specific Cases, Genes, and/or Mutations on the left of this page and then can visualize these results on the right. The Gene/Mutation data for these visualizations comes from the Open-Access MAF files on the GDC Data Portal. There is also a Clinical tab with filters that apply specifically to clinical data.

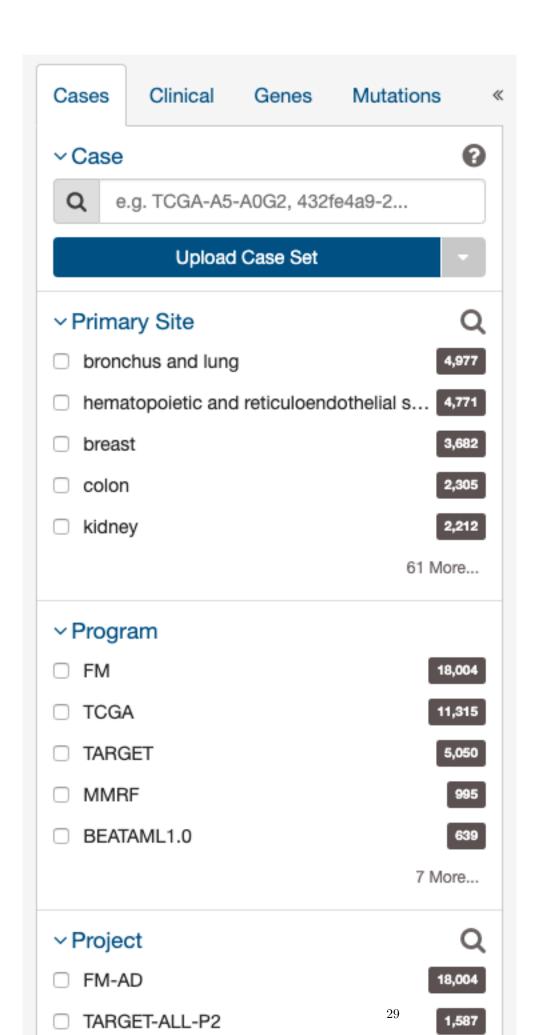


Filters / Facets

On the left of this page, users can create advanced filters to narrow down results to create synthetic cohorts.

Case Filters

The first tab of filters is for cases in the GDC.



These criteria limit the results only to specific cases within the GDC. The default filters available are:

- Case: Specify individual cases using submitter ID (barcode), UUID, or list of Cases ('Case Set').
- Primary Site: Anatomical site of the cancer under investigation or review.
- Program: A cancer research program, typically consisting of multiple focused projects.
- Project: A cancer research project, typically part of a larger cancer research program.
- Disease Type: Type of cancer studied.
- Experimental Strategy: Experimental strategy used for molecular characterization of the cancer.
- Sample Type: Describes the source of a biospecimen used for a laboratory test.
- Available Variation Data: Indicates the types of genomic variation data that a case has been tested for.

Clinical Filters

The second tab of filters is used to specifically explore clinical data for cases in the GDC.

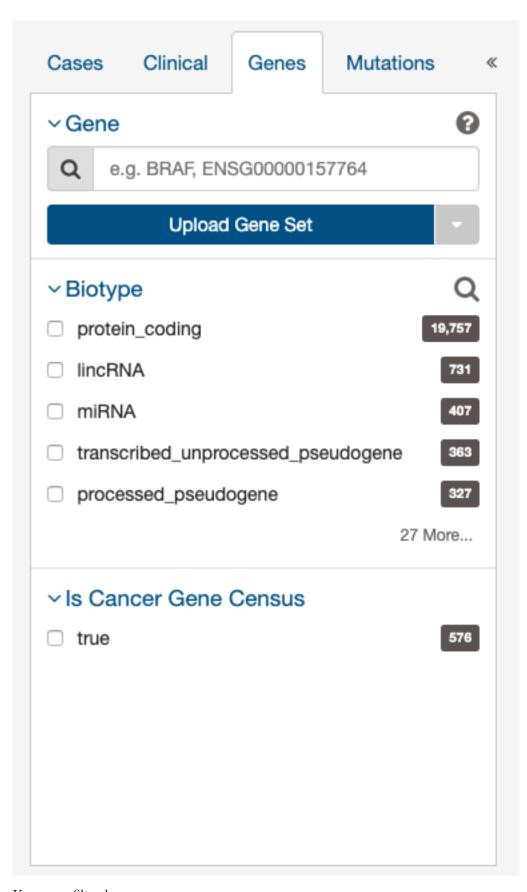
Cases	Clinical	Genes	Mutations	«
Q Se	arch			
Only	show fields	with values (79 fields shown)	
∨ Demo	graphic		Collapse All	
∨ Gen	der			
☐ fem	ale		19,34	1
□ mal	е		17,33	4
□ unk	nown		6	9
□ not	reported			4
□ uns	pecified			1
∨ Race)			
□ not	reported		20,09	6
□ whi	te		13,04	6
□ blac	ck or african	american	1,76	6
□ asia	ın		90	1
□ unk	nown		77:	3
			4 More.	
∨ Ethn	icity			
□ not	reported		31 21,58	8

Users can filter by specific clinical variables, grouped into these categories:

- **Demographic:** Data for the characterization of the patient by means of segmenting the population (e.g. characterization by age, sex, race, etc.).
- **Diagnoses:** Data from the investigation, analysis, and recognition of the presence and nature of disease, condition, or injury from expressed signs and symptoms; also, the scientific determination of any kind; the concise results of such an investigation.
- **Treatments:** Records of the administration and intention of therapeutic agents provided to a patient to alter the course of a pathologic process.
- Exposures: Clinically-relevant patient information not immediately resulting from genetic predispositions.

Gene Filters

The third tab of filters is for genes affected by mutations in the GDC.



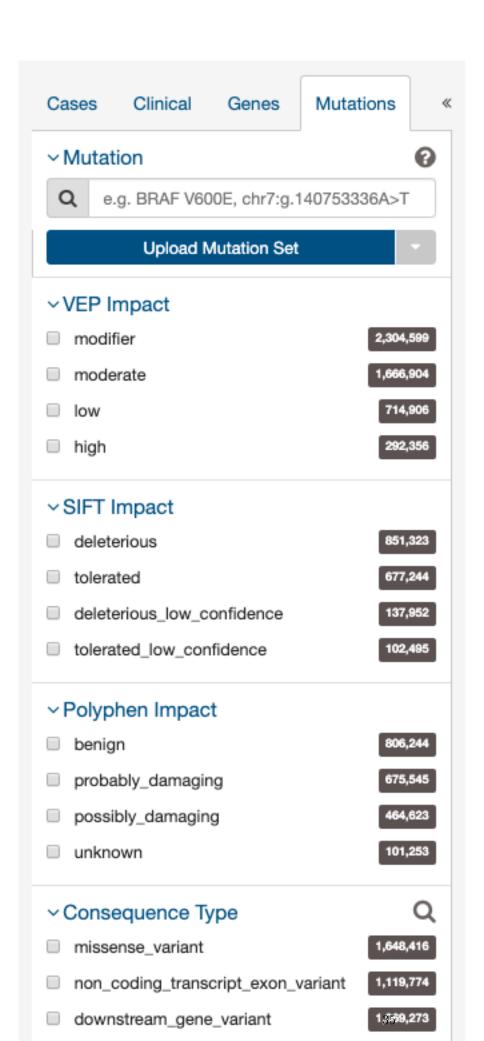
Users can filter by:

• Gene: Specify a Gene Symbol, ID, or list of Genes ('Gene Set').

- Biotype: Classification of the type of gene according to Ensembl. The biotypes can be grouped into protein coding, pseudogene, long noncoding and short noncoding. Examples of biotypes in each group are as follows:
 - Protein coding: IGC gene, IGD gene, IG gene, IGJ gene, IGLV gene, IGM gene, IGV gene, IGZ gene, nonsense
 mediated decay, nontranslating CDS, non stop decay, polymorphic pseudogene, TRC gene, TRD gene, TRJ gene, TRV
 gene.
 - Pseudogene: disrupted domain, IGC pseudogene, IGJ pseudogene, IG pseudogene, IGV pseudogene, processed pseudogene, transcribed processed pseudogene, transcribed unitary pseudogene, transcribed unprocessed pseudogene, translated processed pseudogene, translated unprocessed pseudogene, TRJ pseudogene, TRV pseudogene, unprocessed pseudogene.
 - Long noncoding: 3 prime overlapping ncrna, ambiguous orf, antisense, antisense RNA, lincRNA, macro lincRNA, ncrna host, processed transcript, sense intronic, sense overlapping.
 - Short noncoding: miRNA, miRNA pseudogene, miscRNA, miscRNA pseudogene, Mt rRNA, Mt tRNA, rRNA, scRNA, snlRNA, snoRNA, snRNA, tRNA, tRNA pseudogene, vaultRNA.
- Is Cancer Gene Census: Whether or not a gene is part of The Cancer Gene Census.

Mutation Filters

The final tab of filters is for specific mutations.



Users can filter by:

- Mutation: Unique ID for that mutation. Users can use the following:
 - UUID c7c0aeaa-29ed-5a30-a9b6-395ba4133c63
 - DNA Change chr12:g.121804752delC
 - COSMIC ID COSM202522
 - List of any mutation UUIDs or DNA Change id's ('Mutation Set').
- Impact: A subjective classification of the severity of the variant consequence. These scores are determined using the three following tools:

- Ensembl VEP:

- * HIGH (H): The variant is assumed to have high (disruptive) impact in the protein, probably causing protein truncation, loss of function or triggering nonsense mediated decay.
- * MODERATE (M): A non-disruptive variant that might change protein effectiveness.
- * LOW (L): Assumed to be mostly harmless or unlikely to change protein behavior.
- * MODIFIER (MO): Usually non-coding variants or variants affecting non-coding genes, where predictions are difficult or there is no evidence of impact.

- PolyPhen:

- * probably damaging (PR): It is with high confidence supposed to affect protein function or structure.
- * possibly damaging (PO): It is supposed to affect protein function or structure.
- * benign (BE): Most likely lacking any phenotypic effect.
- * unknown (UN): When in some rare cases, the lack of data does not allow PolyPhen to make a prediction.

- **SIFT**:

- * tolerated: Not likely to have a phenotypic effect.
- * tolerated_low_confidence: More likely to have a phenotypic effect than 'tolerated'.
- * **deleterious:** Likely to have a phenotypic effect.
- * deleterious low confidence: Less likely to have a phenotypic effect than 'deleterious'.
- Consequence Type: Consequence type of this variation; sequence ontology terms.
- Type: A general classification of the mutation.
- Variant Caller: The variant caller used to identify the mutation.
- **COSMIC ID:** This option will filter out only mutations with a COSMIC ID.
- dbSNP rs ID: This option will filter out only mutations with a SNP identifier maintained in dbSNP.

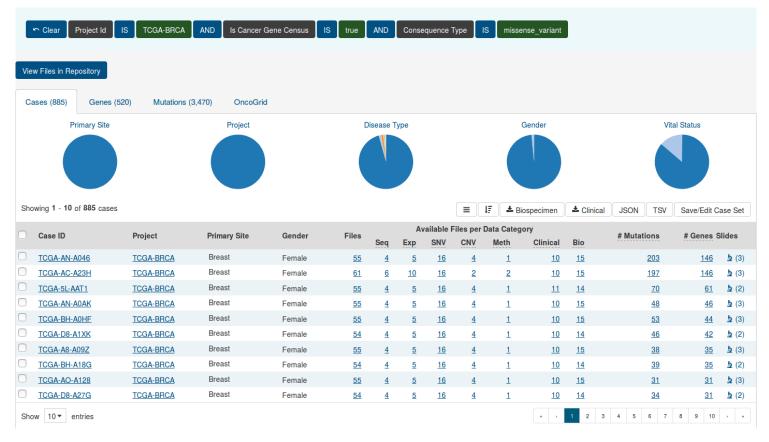
Results

As users add filters to the data on the Exploration Page, the Results section will automatically be updated. Results are divided into different tabs: Cases, Genes, Mutations, and OncoGrid.

To illustrate these tabs, Case, Gene, and Mutation filters have been chosen (Genes in the Cancer Gene Census, that have a missense variant for the TCGA-BRCA project) and a description of what each tab displays follows.

Cases

The Cases tab gives an overview of all the cases/patients who correspond to the filters chosen (Cohort).



The top of this section contains a few pie graphs with categorical information regarding the Primary Site, Project, Disease Type, Gender, and Vital Status.

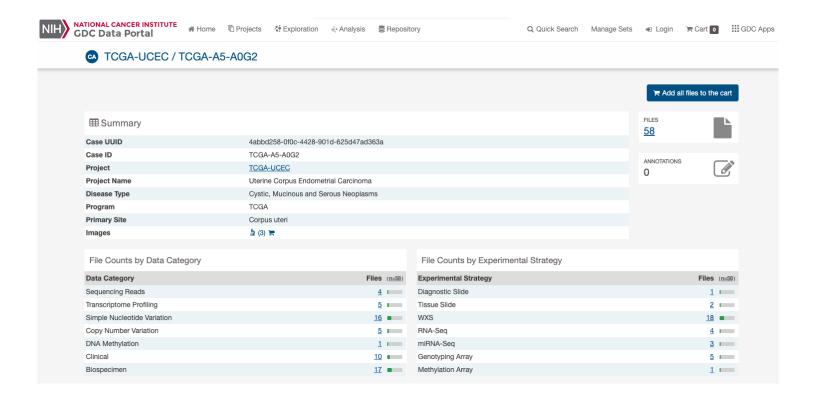
Below these pie charts is a tabular view of cases, which can be exported, sorted and saved using the buttons on the right and includes the following information:

- Case ID (Submitter ID): The Case ID / submitter ID of that case/patient (i.e. TCGA Barcode).
- **Project:** The study name for the project for which the case belongs.
- **Primary Site:** The primary site of the cancer/project.
- Gender: The gender of the case.
- Files: The total number of files available for that case.
- Available Files per Data Category: Seven columns displaying the number of files available in each of the seven data categories. These link to the files for the specific case.
- Mutations: The number of SSMs (simple somatic mutations) detected in that case.
- ___ Genes:___ The number of genes affected by mutations in that case.
- Slides: The total number of slides available for that case. For more information about slide images.

Note: By default, the UUID is not displayed on summary page tables. You can display the UUID by clicking on the icon with 3 parallel lines and checking the UUID option.

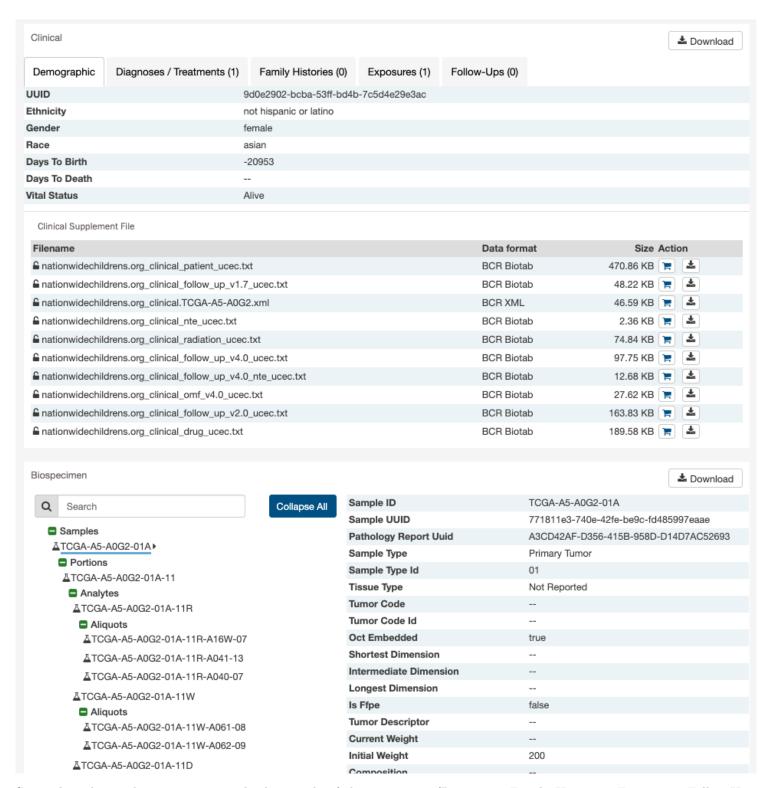
Case Summary Page

The Case Summary Page displays case details including the project and disease information, data files that are available for that case, and the experimental strategies employed. A button in the top-right corner of the page allows the user to add all files associated with the case to the file cart.

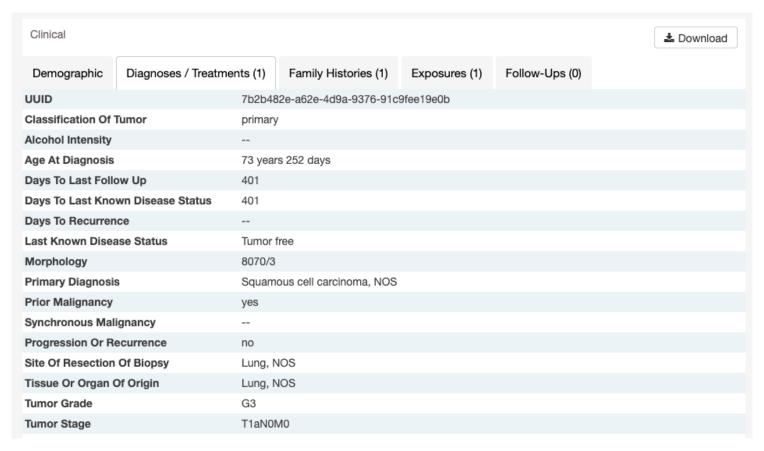


Clinical and Biospecimen Information

The page also provides clinical and biospecimen information about that case. Links to export clinical and biospecimen information in JSON format are provided.



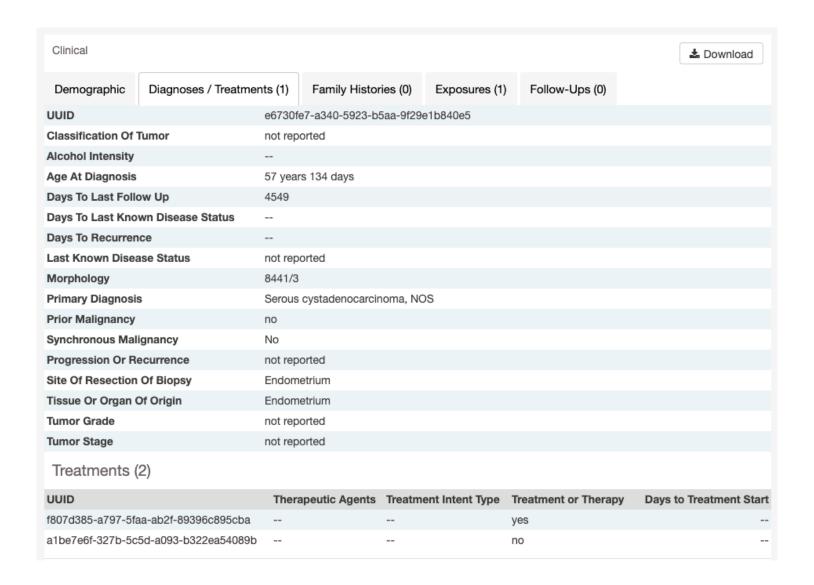
Some clinical records can support multiple records of the same type (Diagnoses, Family Histories, Exposures, Follow-Ups, Molecular Tests). If only one record exists, the UUID of the record is provided at the top of the corresponding tab.



If there are multiple records, they are listed as horizontal tabs.

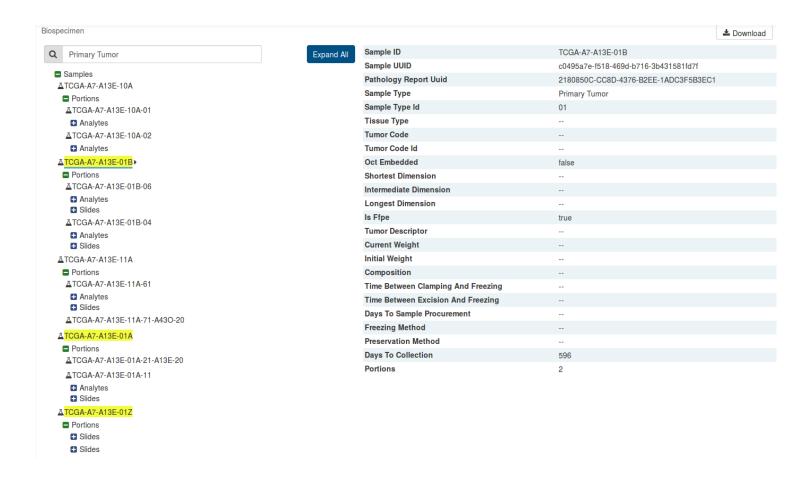
Clinical								≛ Download
Demographic	Diagnoses / Treatments (1)	Family I	Histories (1)	Exposure	es (0) Follo	w-Ups (10)		
01fa4de9	UUID		01fa4de9-6c9	91-4c5c-be	26-739a22dc2	9fb		
onaques	Days To Follow Up		751					
5fa2e435	Comorbidity							
	Risk Factor							
	Progression Or Recurrence Type							
a45e67ef	Progression Or Recurrence							
	Disease Response							
d050102b	ВМІ							
	Height							
8fa6cefb	Weight							
	Ecog Performance Status							
	Karnofsky Performance Status							
03b92e6a	Progression Or Recurrence Anat	omic Site	e					
	Reflux Treatment Type							
95faf2c6	Molecular Tests (15)							
e038a4d0	UUID		Molecular Analysis Method	Gene Symbol	Laboratory Te	st Bio	specimen e	Test Value
4cd0a0a4	96642d4d-9a59-420d-bd80-61ad2	22924871	Not Reported	Not Reported	M Protein	Blo	od	
fd8249ef	23865352-2109-4706-bfb6-fd574a	c36a3e	Not Reported	Not Reported	Albumin	Blo	od	44 g/L
10024901	165c006e-91a9-40cf-b39b-8b4408	Ba2a060	Not Reported	Not Reported	Platelets	Blo	od	183 count x10^9
	e5e393b5-14d3-4c87-a80b-8e519	4e94572	Not Reported	Not Reported	Lactate Dehydrogenas	е		3.334 ukat/L
	ccd8a5f6-248b-460b-b237-46ed3	5428f93	Not Reported	Not Reported	Creatinine	Blo	od	72 umol/L
	58325afb-f701-4f9e-9d4a-0155ed0	016c11	Not Reported	Not Reported	Beta 2 Microgl	obulin Blo	od	3.32 mcg/mL

Some record types are further nested under another. For example, a Diagnosis record may have multiple associated Treatment records. Or a Follow-Up record may have multiple associated Molecular Test Records. The associated sub-records are listed in a table on the tab.



Biospecimen Search

A search filter just below the biospecimen section can be used to find and filter biospecimen data. The wildcard search will highlight entities in the tree that match the characters typed. This will search both the case submitter ID, as well as the additional metadata for each entity. For example, searching 'Primary Tumor' will highlight samples that match that type.



Most Frequent Somatic Mutations for a Case

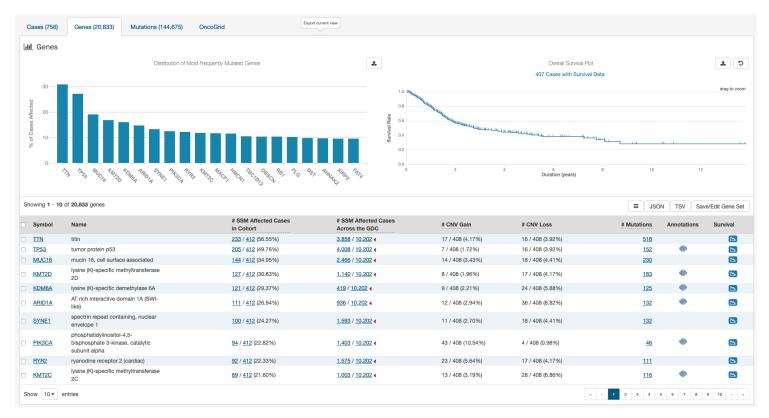
The Case Entity Page also lists the mutations found in that particular case.



For more information, please go to the Most Frequent Somatic Mutation section.

Genes

The Genes tab will give an overview of all the genes that match the criteria of the filters (Cohort).



The top of this tab contains a bar graph of the most frequently mutated genes. Hovering over each bar in the plot will display information about the percentage of cases affected. In addition, this section contains a survival curve. The survival curve is calculated using the Kaplan-Meier estimator based on all cases with survival data within the specified Exploration Page search. For more information on how these values are determined, please go to the Survival Analysis section. Users may choose to download the underlying data in JSON or TSV format or an image of the graph in SVG or PNG format by clicking the download icon at the top of each graph.

Below these graphs is a tabular view of the genes affected, which includes the following information:

- Symbol: The gene symbol, which links to the Gene Summary Page.
- Name: Full name of the gene.
- ___ SSM Affected Cases in Cohort:___ The number of cases affected by SSMs (simple somatic mutations) in the Cohort.
- ___ SSM Affected Cases Across the GDC:___ The number of cases within all the projects in the GDC that contain a mutation on this gene. Clicking the red arrow will display the cases broken down by project.
- ___ CNV Gain:___ The number of CNV (copy number variation) events detected in that gene which resulted in an increase (gain) in the gene's copy number.
- ___ CNV Loss:___ The number of CNV events detected in that gene which resulted in a decrease (loss) in the gene's copy number.
- Mutations: The number of SSMs (simple somatic mutations) detected in that gene.
- Annotations: Includes a COSMIC symbol if the gene belongs to The Cancer Gene Census.
- Survival: An icon that, when clicked, will plot the survival rate between cases in the project with mutated and non-mutated forms of the gene.

Gene Summary Page

Gene Summary Pages describe each gene with mutation data and provides results related to the analyses that are performed on these genes.

The summary section of the Gene Page contains the following information:

■ Summary		External Reference	es
Symbol	TP53	NCBI Gene	♂ <u>7157</u>
Name	tumor protein p53	UniProtKB Swiss-Prot	☑ P04637
Synonyms	LFS1	HGNC	☑ HGNC:11998
	p53	OMIM	♂ <u>191170</u>
Туре	protein_coding	Ensembl	☑ ENSG00000141510
Location	chr17:7661779-7687550 (GRCh38)	CIVIC	₫ <u>45</u>
Strand	-		
Description	This gene encodes a tumor suppressor protein containing transcriptional activation, DNA binding, and oligomerization domains. The encoded protein responds to diverse cellular stresses to regulate expression of target genes, thereby inducing cell cycl		
Annotation	Cancer Gene Census		

• Symbol: The gene symbol.

• Name: Full name of the gene.

• Synonyms: Synonyms of the gene name or symbol, if available.

• Type: A broad classification of the gene.

• Location: The chromosome on which the gene is located and its coordinates.

• Strand: If the gene is located on the forward (+) or reverse (-) strand.

• Description: A description of gene function and downstream consequences of gene alteration.

• Annotation: A notation/link that states whether the gene is part of The Cancer Gene Census.

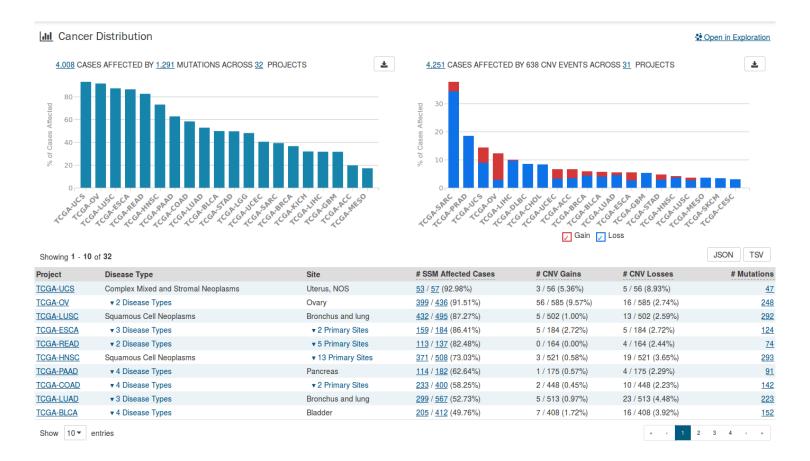
External References

A list with links that lead to external databases with additional information about each gene is displayed here. These external databases include:

- Entrez
- Uniprot
- Hugo Gene Nomenclature Committee
- Online Mendelian Inheritance in Man
- Ensembl
- CIViC

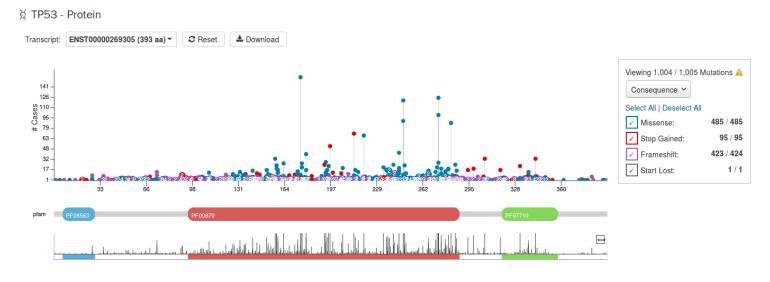
Cancer Distribution

A table and two bar graphs show how many cases are affected by mutations and copy number variation within the gene as a ratio and percentage. Each row/bar represents the number of cases for each project. The final column in the table lists the number of unique mutations observed on the gene for each project.



Protein Viewer

Mutations and their frequency across cases are mapped to a graphical visualization of protein-coding regions with a lollipop plot. Pfam domains are highlighted along the x-axis to assign functionality to specific protein-coding regions. The bottom track represents a view of the full gene length. Different transcripts can be selected by using the drop-down menu above the plot.



The panel to the right of the plot allows the plot to be filtered by mutation consequences or impact. The plot will dynamically change as filters are applied. Mutation consequence and impact is denoted in the plot by color.

Note: The impact filter on this panel will not display the annotations for alternate transcripts.

The plot can be viewed at different zoom levels by clicking and dragging across the x-axis, clicking and dragging across the bottom track, or double clicking the pfam domain IDs. The Reset button can be used to bring the zoom level back to its original

position. The plot can also be exported as a PNG image, SVG image or as JSON formatted text by choosing the Download button above the plot.

Most Frequent Somatic Mutations

The 20 most frequent mutations in the gene are displayed as a bar graph that indicates the number of cases that share each mutation.



A table is displayed below that lists information about each mutation including:

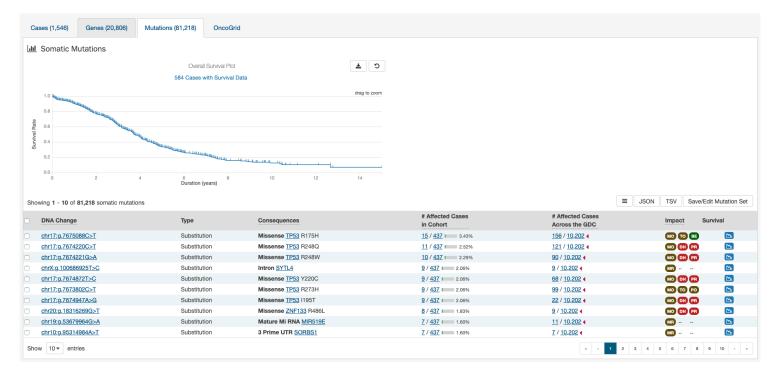
- DNA Change: The chromosome and starting coordinates of the mutation are displayed along with the nucleotide differences between the reference and tumor allele.
- Type: A general classification of the mutation.
- Consequences: The effects the mutation has on the gene coding for a protein (i.e. synonymous, missense, non-coding transcript).
- ___ Affected Cases in Gene:___ The number of affected cases, expressed as number across all mutations within the selected Gene.
- ___ Affected Cases Across GDC:___ The number of affected cases, expressed as number across all projects. Choosing the arrow next to the percentage will expand the selection with a breakdown of each affected project.
- Impact: A subjective classification of the severity of the variant consequence. This is determined by three different tools:
 - Ensembl VEP
 - PolyPhen
 - SIFT

Clicking the Open in Exploration button will navigate the user to the Exploration Page, showing the same results in the table (mutations filtered by the gene).

Mutations

The Mutations tab will give an overview of all the mutations that match the criteria of the filters (Cohort).

Open-access mutation data is displayed by defualt. To access controlled access mutations, users must apply to the correct data access authority, be granted access, and login to the portal. If a user is logged in and has been granted access to controlled-access mutations, they will be integrated with open-access mutations throughout the portal visualizations and counts.



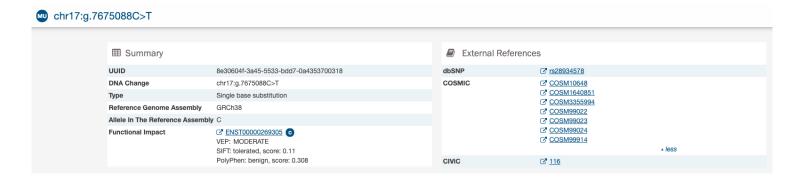
At the top of this tab contains a survival curve. The survival curve is calculated using the Kaplan-Meier estimator based on all cases with survival data within the specified Exploration Page search. For more information on how these values are determined, please go to the Survival Analysis section. Users may choose to download the underlying data in JSON or TSV format or an image of the graph in SVG or PNG format by clicking the download icon at the top of the graph.

A table is displayed below that lists information about each mutation:

- **DNA Change:** The chromosome and starting coordinates of the mutation are displayed along with the nucleotide differences between the reference and tumor allele.
- Type: A general classification of the mutation.
- Consequences: The effects the mutation has on the gene coding for a protein (i.e. synonymous, missense, non-coding transcript). A link to the Gene Summary Page for the gene affected by the mutation is included.
- Affected Cases in Cohort: The number of affected cases in the Cohort as a fraction and as a percentage.
- ___ Affected Cases in Across all Projects:___ The number of affected cases, expressed as number across all projects. Clicking the arrow next to the percentage will display a breakdown of each affected project.
- Impact: A subjective classification of the severity of the variant consequence. This is determined by three different tools:
 - Ensembl VEP
 - PolyPhen
 - SIFT
- Survival: An icon that when clicked, will plot the survival rate between the gene's mutated and non-mutated cases.

Mutation Summary Page

The Mutation Summary Page contains information about one somatic mutation and how it affects the associated gene. Each mutation is identified by its chromosomal position and nucleotide-level change.



- UUID: A unique identifier (UUID) for this mutation.
- DNA Change: Denotes the chromosome number, position, and nucleotide change of the mutation.
- Type: A broad categorization of the mutation.
- Reference Genome Assembly: The reference genome in which the chromosomal position refers to.
- Allele in the Reference Assembly: The nucleotide(s) that compose the site in the reference assembly.
- Functional Impact: A subjective classification of the severity of the variant consequence.

External References

A separate panel contains links to databases that contain information about the specific mutation. These include dbSNP, COSMIC, and CIViC.

Consequences

The consequences of the mutation are displayed in a table. The set of consequence terms, defined by the Sequence Ontology.

EXECUTE Consequences

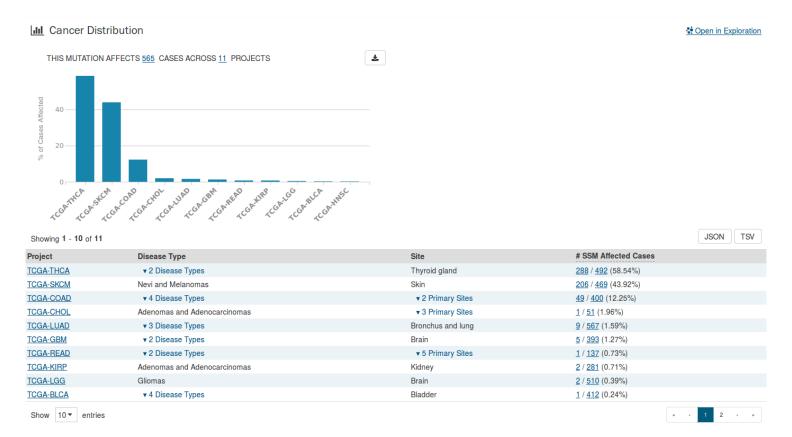


The fields that describe each consequence are listed below:

- Gene: The symbol for the affected gene.
- AA Change: Details on the amino acid change, including compounds and position, if applicable.
- Consequence: The biological consequence of each mutation.
- Coding DNA Change: The specific nucleotide change and position of the mutation within the gene.
- Impact: A subjective classification of the severity of the variant consequence. This is determined by three different tools:
 - Ensembl VEP
 - PolyPhen
 - SIFT
- Strand: If the gene is located on the forward (+) or reverse (-) strand.
- Transcript(s): The transcript(s) affected by the mutation. Each contains a link to the Ensemble entry for the transcript.

Cancer Distribution

A table and bar graph shows how many cases are affected by the particular mutation. Each row/bar represents the number of cases for each project.



The table contains the following fields:

- Project ID: The ID for a specific project.
- **Disease Type**: The disease associated with the project.
- Site: The anatomical site affected by the disease.
- ___ SSM Affected Cases___: The number of affected cases and total number of cases displayed as a fraction and percentage.

Protein Viewer

The protein viewer displays a plot representing the position of mutations along the polypeptide chain. The y-axis represents the number of cases that exhibit each mutation, whereas the x-axis represents the polypeptide chain sequence. Pfam domains that were identified along the polypeptide chain are identified with colored rectangles labeled with pfam IDs. See the Gene Summary Page for additional details about the protein viewer.



OncoGrid

The Exploration Page includes an OncoGrid plot of the cases with the most mutations, for the top 50 mutated genes affected by high impact mutations. Genes displayed on the left of the grid (Y-axis) correspond to individual cases on the bottom of the grid (X-axis). Additionally, the plot also indicates in each cell any CNV events detected for these top mutated cases and genes.



The grid is color-coded with a legend at the top which describes what type of mutation consequence and CNV event is observed for each gene/case combination. Clinical information and the available data for each case are available at the bottom of the grid.

The right side of the grid displays additional information about the genes:

- Gene Sets: Describes whether a gene is part of The Cancer Gene Census. (The Cancer Gene Census is an ongoing effort to catalogue those genes for which mutations have been causally implicated in cancer)
- ___ Cases Affected:___ Identifies all cases in the GDC affected with a mutation in this gene

OncoGrid Options

To facilitate readability and comparisons, drag-and-drop can be used to reorder the gene rows. Double clicking a row in the "Cases Affected" bar at the right side of the graphic launches the respective Gene Summary Page. Hovering over a cell will display information about the mutation such as its ID, affected case, and biological consequence. Clicking on the cell will bring the user to the respective Mutation Summary Page.

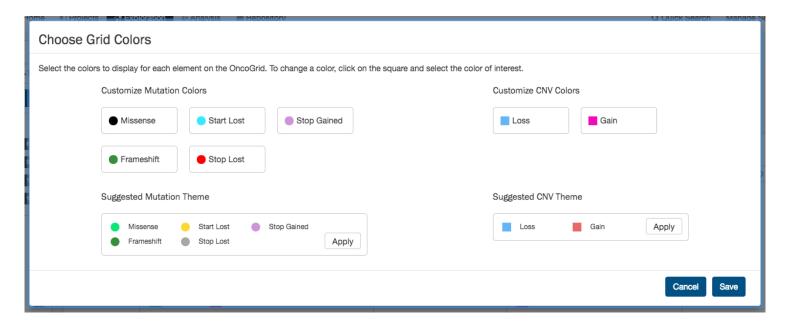
A tool bar at the top right of the graphic allows the user to export the data as a JSON object, PNG image, or SVG image. Seven buttons are available in this toolbar:

- Customize Colors: Users can customize the colors that represent mutation consequence types and CNV gains/losses.
- **Download:** Users can choose to export the contents either to a static image file (PNG or SVG format) or the underlying data in JSON format.
- Reload Grid: Sets all OncoGrid rows, columns, and zoom levels back to their initial positions.
- Cluster Data: Clusters the rows and columns to place mutated genes with the same cases and cases with the same mutated genes together.
- Toggle Heatmap: The view can be toggled between cells representing mutation consequences or number of mutations in each gene.
- Toggle Gridlines: Turn the gridlines on and off.
- Toggle Crosshairs: Turns crosshairs on, so that users can zoom into specific sections of the OncoGrid.
- Fullscreen: Turns Fullscreen mode on/off.

OncoGrid Color Picker

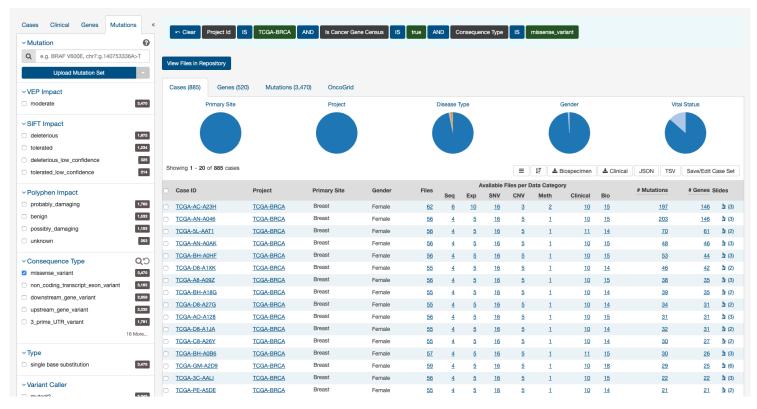
To customize the colors for mutation consequence types and CNV gains/losses, a user can click the color picker icon in the OncoGrid toolbar.

- Customize Colors: Opens a control where the user can pick their own colors or apply a suggested theme and save their changes.
- Reset to Default: Resets all colors to the defaults initially used by OncoGrid.

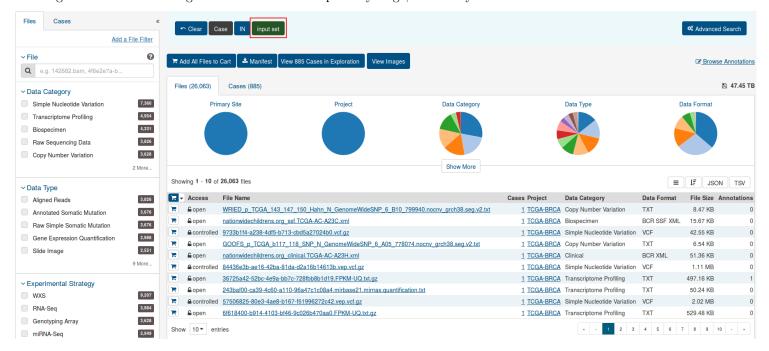


File Navigation

After utilizing the Exploration Page to narrow down a specific cohort, users can find the specific files that relate to this group by clicking on the View Files in Repository button as shown in the image below.



Clicking this button will navigate the users to the Repository Page, filtered by the cases within the cohort.



The filters chosen on the Exploration Page are displayed as an input set on the Repository Page. Additional filters may be added on top of this input set, but the original set cannot be modified and instead a new input set must be created from original data.

Survival Analysis

The survival analysis, which is seen in both the Gene and Mutation tabs, is used to analyze the occurrence of event data over time. In the GDC, survival analysis is performed on the mortality of the cases. Thus, the values are retrieved from GDC Data Dictionary properties and a survival analysis requires the following fields:

- Data on the time to a particular event (days to death or last follow up).
 - Fields: demographic.days to death or demographic.days to last follow up
- Information on whether the event has occurred (alive/deceased).
 - Fields: demographic.vital_status
- Data split into different categories or groups (i.e. gender, etc.).
 - Fields: demographic.gender

The survival analysis in the GDC uses a Kaplan-Meier estimator:

$$\hat{S}(t) = \prod_{i:t_i \le t} \left(1 - \frac{d_i}{n_i} \right)$$

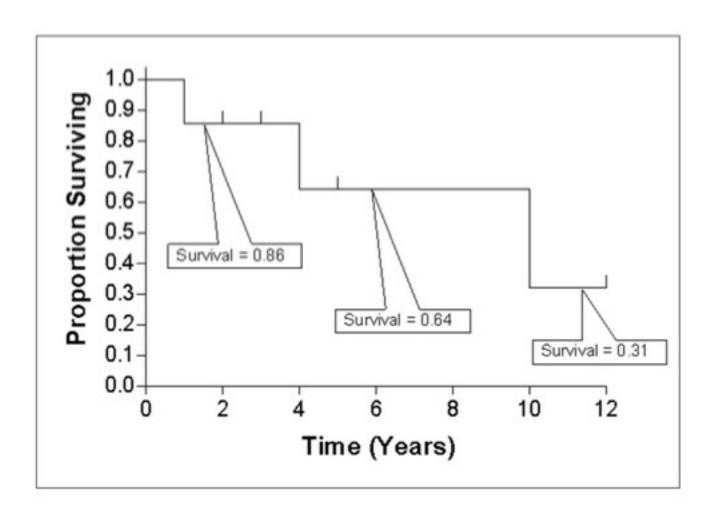
Where:

- S(t) is the estimated survival probability for any particular one of the t time periods.
- ni is the number of subjects at risk at the beginning of time period ti.
- and di is the number of subjects who die during time period ti.

The table below is an example data set to calculate survival for a set of seven cases:

overall_survial_time (Years)	interval		# of donors at	# of censored	# of donors at	# of donors	estimated interval	estimated cumulative	
	start	end	risk at start of interval (r)	donors during interval (c)	risk at end of interval (n=r-c)	died at end of interval (d)	survival probability ((n-d)/n)	survival probability at end of interva	
0	0							1	
1	0	1	7	0	7	1	(7-1)/7 = 0.86	1 * 0.86 = 0.86	
4	1	4	6	2	4	1	(4-1)/4 = 0.75	0.86 * 0.75 = 0.64	
10	4	10	3	1	2	1	(2-1)/2 = 0.5	0.86 * 0.75 * 0.5 = 0.31	
>12	10	12	1	0	1	0	(1-0)/1 = 1.0	0.86 * 0.75 * 0.5 * 1.0 = 0.3 1	

The calculated cumulated survival probability can be plotted against the interval to obtain a survival plot like the one shown below.



Chapter 4

Repository

Repository

The Repository Page is the primary method of accessing data in the GDC Data Portal. It provides an overview of all cases and files available in the GDC and offers users a variety of filters for identifying and browsing cases and files of interest. Users can access the Repository Page from the GDC Data Portal Home Page or from the Data Portal toolbar.

Filters / Facets

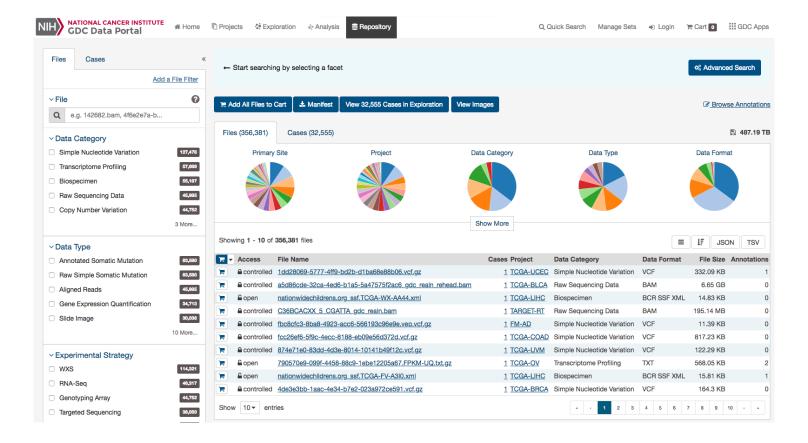
On the left, a panel of data facets allows users to filter cases and files using a variety of criteria. If facet filters are applied, the tabs on the right will display information about matching cases and files. If no filters are applied, the tabs on the right will display information about all available data.

On the right, two tabs contain information about available data:

- Files tab provides a list of files, select information about each file, and links to individual file detail pages.
- Cases tab provides a list of cases, select information about each case, and links to individual case summary pages.

The banner above the tabs on the right displays any active facet filters and provides access to advanced search.

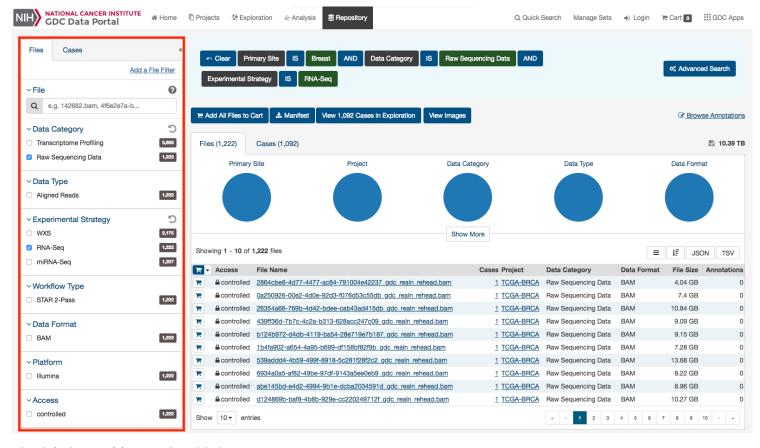
The top of the Repository Page, in the "Files" tab, contains a few summary pie charts for Primary Sites, Projects, Data Category, Data Type, and Data Format. These reflect all available data or, if facet filters are applied, only the data that matches the filters. Clicking on a specific slice in a pie chart, or on a number in a table, applies corresponding facet filters. The scope of these pie chart will change depending on whether you have the "Files" tab or the "Cases" tab selected.



Facets Panel

Facets represent properties of the data that can be used for filtering. The facets panel on the left allows users to filter the cases and files presented in the tabs on the right.

The facets panel is divided into two tabs, with the Files tab containing facets pertaining to data files and experimental strategies, while the Cases tab containing facets pertaining to the cases and biospecimen information. Users can apply filters in both tabs simultaneously. The applied filters will be displayed in the banner above the tabs on the right, with the option to open the filter in Advanced Search to further refine the query.



The default set of facets is listed below.

Files facets tab:

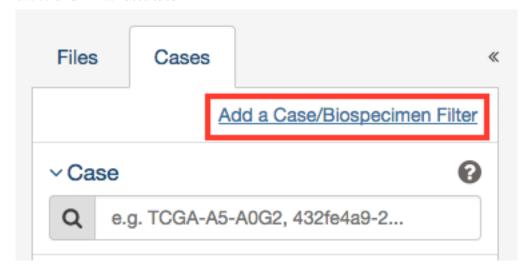
- File: Specify individual files using filename or UUID.
- Data Category: A high-level data file category, such as "Raw Sequencing Data" or "Transcriptome Profiling".
- Data Type: Data file type, such as "Aligned Reads" or "Gene Expression Quantification". Data Type is more granular than Data Category.
- Experimental Strategy: Experimental strategies used for molecular characterization of the cancer.
- Workflow Type: Bioinformatics workflow used to generate or harmonize the data file.
- Data Format: Format of the data file.
- Platform: Technological platform on which experimental data was produced.
- Access Level: Indicator of whether access to the data file is open or controlled.

Cases facets tab:

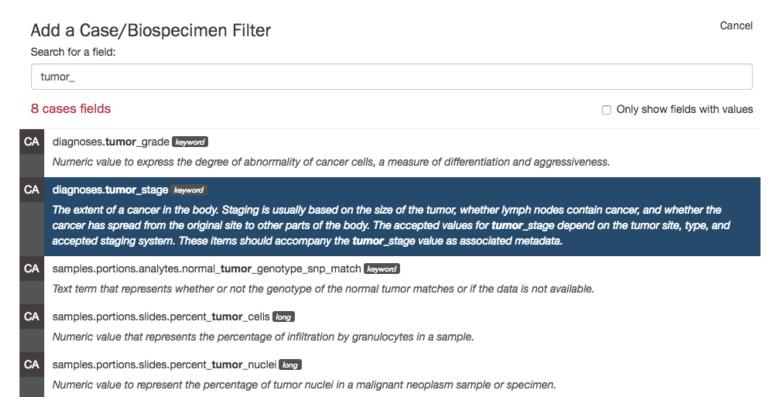
- Case: Specify individual cases using submitter ID (barcode) or UUID.
- Case ID: Search for cases using a part (prefix) of the submitter ID (barcode).
- Primary Site: Anatomical site of the cancer under investigation or review.
- Program: A cancer research program, typically consisting of multiple focused projects.
- Project: A cancer research project, typically part of a larger cancer research program.
- Disease Type: Type of cancer studied.
- **Gender**: Gender of the patient.
- Age at Diagnosis: Patient age at the time of diagnosis.
- Vital Status: Indicator of whether the patient was living or deceased at the date of last contact.
- Days to Death: Number of days from date of diagnosis to death of the patient.
- Race: Race of the patient.
- Ethnicity: Ethnicity of the patient.

Adding Custom Facets

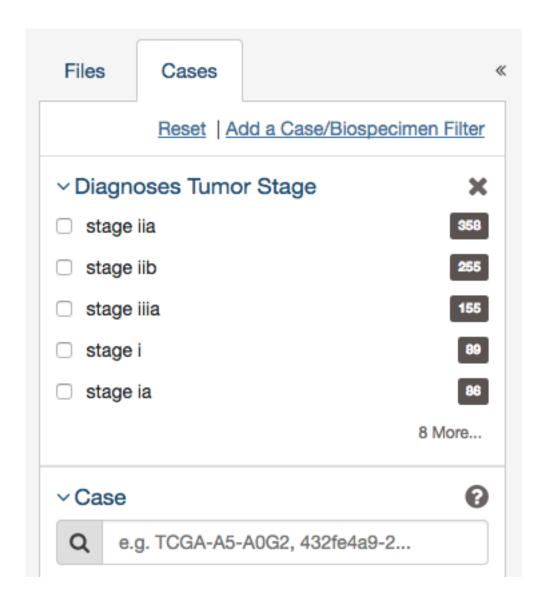
The Repository Page provides access to additional data facets beyond the automatically listed group filters. Facets corresponding to additional properties listed in the GDC Data Dictionary can be added using the "Add a Filter" link available at the top of the Cases and Files facet tabs:



The link opens a search window that allows the user to find an additional facet by name or description. Not all facets have values available for filtering; checking the "Only show fields with values" checkbox will limit the search results to only those that do. When selecting a facet from the list of search results below the search box will add it to the facets panel.

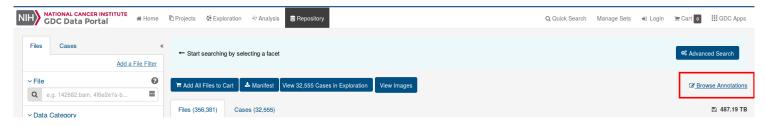


Newly added facets will show up at the top of the facets panel and can be removed individually by clicking on the " \mathbf{x} " to the right of the facet name. The default set of facets can be restored by clicking "Reset".

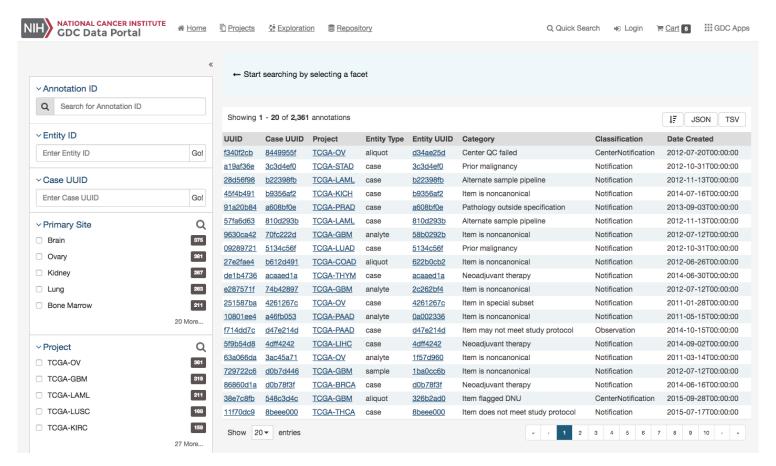


Annotations View

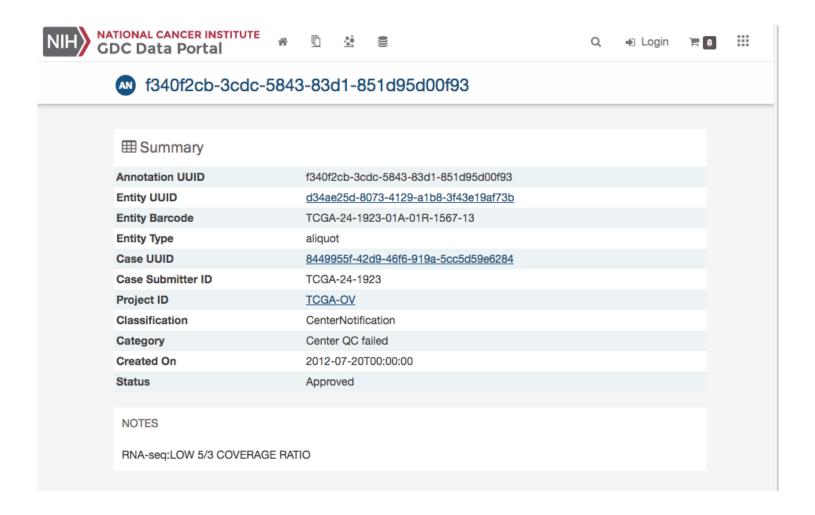
The Annotations View provides an overview of the available annotations and allows users to browse and filter the annotations based on a number of annotation properties (facets), such as the type of entity the annotation is attached to or the annotation category. This page can be found by clicking on the Browse Annotations link, located at the top right of the repository page.



The view presents a list of annotations in tabular format on the right, and a facet panel on the left that allows users to filter the annotations displayed in the table. If facet filters are applied, the tabs on the right will display only the matching annotations. If no filters are applied, the tabs on the right will display information about all available annotations.



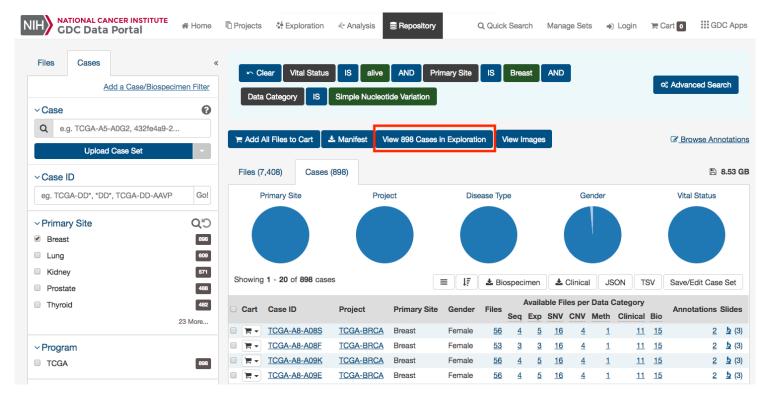
Clicking on an annotation ID in the annotations list will take the user to the Annotation Summary Page. The Annotation Summary Page provides more details about a specific annotation.



Results

Navigation

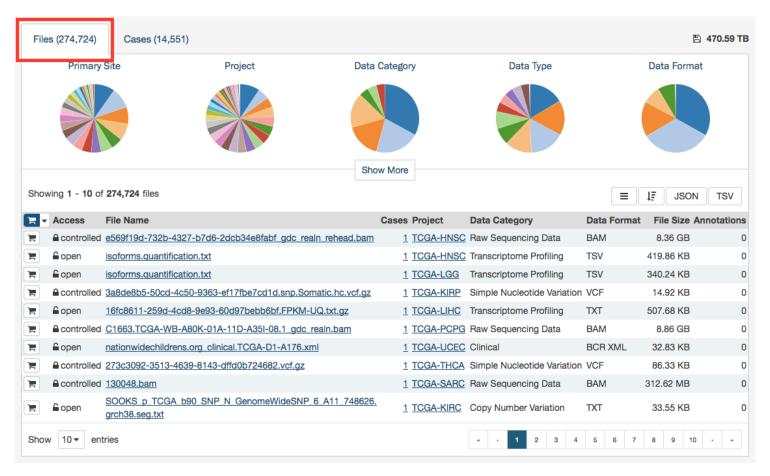
After utilizing the Repository Page to narrow down a specific set of cases, users can choose to continue to explore the mutations and genes affected by these cases by clicking the View Cases in Exploration button as shown in the image below.



Clicking this button will navigate the users to the Exploration Page, filtered by the cases within the cohort.

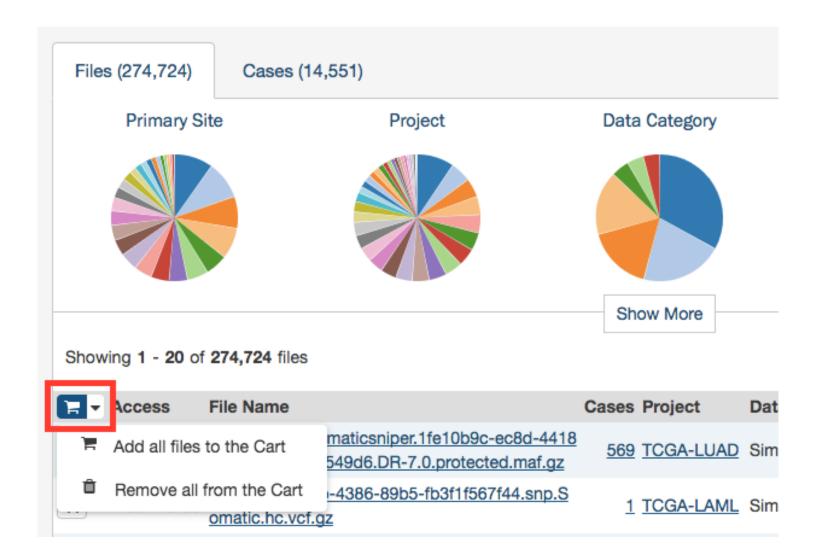
Files List

The Files tab on the right provides a list of available files and select information about each file. If facet filters are applied, the list includes only matching files. Otherwise, the list includes all data files available in the GDC Data Portal.



The "File Name" column includes links to File Summary Pages where the user can learn more about each file.

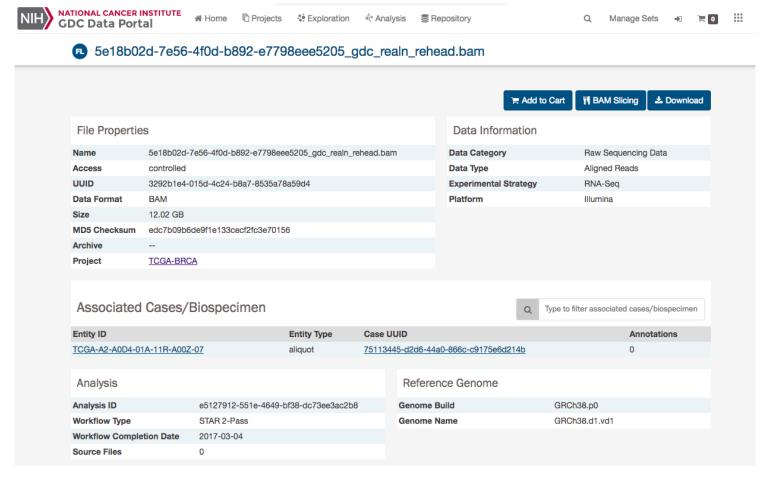
Users can add individual file(s) to the cart using the cart button next to each file. Alternatively, all files that match the current facet filters can be added to the cart using the menu in the top left corner of the table:



File Summary Page

The File Summary page provides information about a data file, including file properties like size, MD5 checksum, and data format; information on the type of data included; links to the associated cases and biospecimen; and information about how the data file was generated or processed.

The page also includes buttons to download the file, add it to the file cart, or (for BAM files) utilize the BAM slicing function.



In the lower section of the screen, the following tables provide more details about the file and its characteristics:

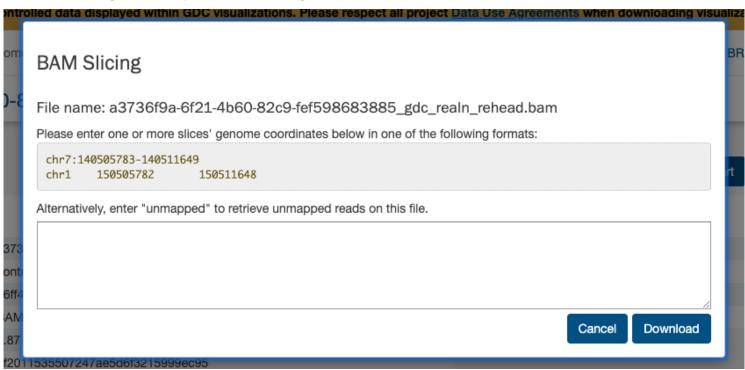
- Associated Cases / Biospecimen: List of cases or biospecimen the file is directly attached to.
- Analysis and Reference Genome: Information on the workflow and reference genome used for file generation.
- Read Groups: Information on the read groups associated with the file.
- Metadata Files: Experiment metadata, run metadata and analysis metadata associated with the file.
- Downstream Analysis Files: List of downstream analysis files generated by the file.
- File Versions: List of all versions of the file.

Read Group	ID	Is Paired End F	Read Length	Library Name	Sequencir	ng Center	S	equencing D	ate	
33e057b9-e4	183-41ee-a837-32a5bf6a1e36	true 5	50	unknown	UNC					
Downstre	eam Analyses Files									
File Name		Data Category	Data Type		Data Format	Analysis	workflow	File Size	Actio	n
8b178cb1-d2	22e-4657-80c6-d7efcddf43a6.htseq.counts.gz	Transcriptome Profili	ing Gene Expressi	on Quantification	TXT	HTSeq -	Counts	255.72 KB		Ł
8b178cb1-d2	22e-4657-80c6-d7efcddf43a6.FPKM-UQ.txt.gz	Transcriptome Profili	ing Gene Expressi	on Quantification	TXT	HTSeq -	FPKM-UQ	521.22 KB	H	7
8b178cb1-d2	22e-4657-80c6-d7efcddf43a6.FPKM.txt.gz	Transcriptome Profili	ing Gene Expressi	on Quantification	TXT	HTSeq -	FPKM	519.72 KB		±
File Vers	ions							± □	Downlo	ad
Version	File UUID				Release Dat	е	Releas	e Number		
	3292b1e4-015d-4c24-b8a7-8535a78a59d4	Current Version			2018-08-08		13.0			

Note: The Legacy Archive will not display the "Workflow, Reference Genome and Read Groups" sections (these sections are applicable to the GDC harmonization pipeline only). However, it may provide information on archives and metadata files like MAGE-TABs and SRA XMLs. For more information, please refer to the section Legacy Archive.

BAM Slicing

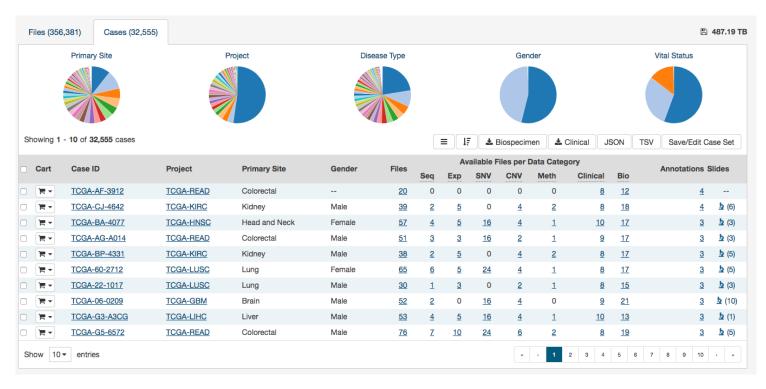
BAM file Summary Pages have a "BAM Slicing" button. This function allows the user to specify a region of a BAM file for download. Clicking on it will open the BAM Slicing window:



During preparation of the slice, the icon on the BAM Slicing button will be spinning, and the file will be offered for download to the user as soon as it is ready.

Cases List

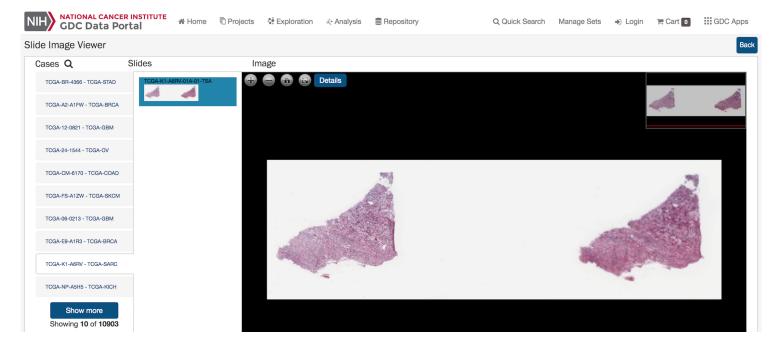
The Cases tab on the right provides a list of available cases and select information about each case. If facet filters are applied, the list includes only matching cases. Otherwise, the list includes all cases available in the GDC Data Portal.



From the left side, the list starts with a shopping cart icon, allowing the user to add all files associated with a case to the file cart for downloading at a later time. The following columns in the list includes links to Case Summary Pages in the Case UUID column, the Submitter ID (i.e. TCGA Barcode), and counts of the available file types for each case. Clicking on a count will apply facet filters to display the corresponding files. On the last column, there are image slide icons and a number that indicate whether there are slide images available and how many.

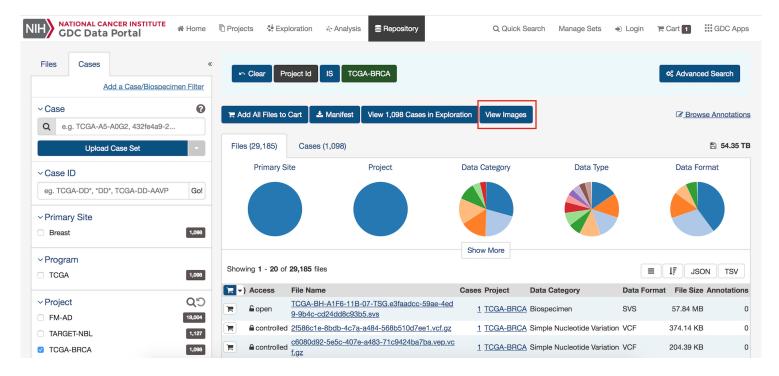
Image Viewer

The Image Viewer allows users to visualize tissue and diagnostic slide images.

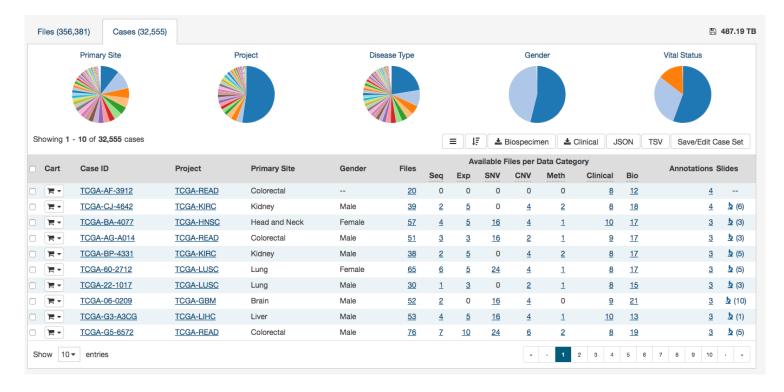


How to Access the Image Viewer

• Repository Page: From the main search on the Repository Page by clicking on the "View images" button. It will display the tissue slide images of all the cases resulting from the query.



• Case Table in Repository Page: Click on the image viewer icon in the Case table. It will display in the image viewer all the tissue slide images attached to the Case.



• Case Summary Page: Selecting a Case ID in the Repository Cases table will direct the user to the Case Summary Page. For cases with images, the Image Viewer icon will appear in the Case Summary section or in the Biospecimen - Slides details section. Clicking on the Image Viewer icon will display the Image Viewer for the slide images attached to the case.

CA TCGA-UCEC / TCGA-A5-A0G2 Add all files to the cart **FILES III** Summary <u>34</u> **Case UUID** 4abbd258-0f0c-4428-901d-625d47ad363a Case ID TCGA-A5-A0G2 **ANNOTATIONS Project TCGA-UCEC** 0 **Project Name** Uterine Corpus Endometrial Carcinoma **Disease Type** Uterine Corpus Endometrial Carcinoma **Program TCGA Primary Site** Uterus **Images ½** (2) 📜 CA TCGA-UCEC / TCGA-A5-A0G2 Biospecimen ▲ Download Slide ID TCGA-A5-A0G2-01A-01-TS1 Q Search Collapse All Slide UUID ac44810a-1366-4bc5-aa98-7d98ef8b6329 Samples 20 **Percent Lymphocyte Infiltration △TCGA-A5-A0G2-10A Percent Tumor Nuclei** 90 ■ Portions

• The Image File Page: You can visualize the slide image directly in the File Summary Page by selecting an image file in the Repository's files table.

Percent Monocyte Infiltration

Percent Neutrophil Infiltration

Percent Granulocyte Infiltration

Percent Inflam Infiltration
Number Proliferating Cells
Percent Tumor Cells

10

5

5

85

P 🗯 🖛

TOP

Percent Normal Cells

Percent Stromal Cells
Percent Eosinophil Infiltration

Section Location

Percent Necrosis

Slide Image

∆TCGA-A5-A0G2-01A

∆TCGA-A5-A0G2-01A-21-A162-20

△TCGA-A5-A0G2-01A-01-TS1▶

△TCGA-A5-A0G2-01A-01-BS1

△TCGA-A5-A0G2-01A-11

Portions

♣ Analytes➡ Slides

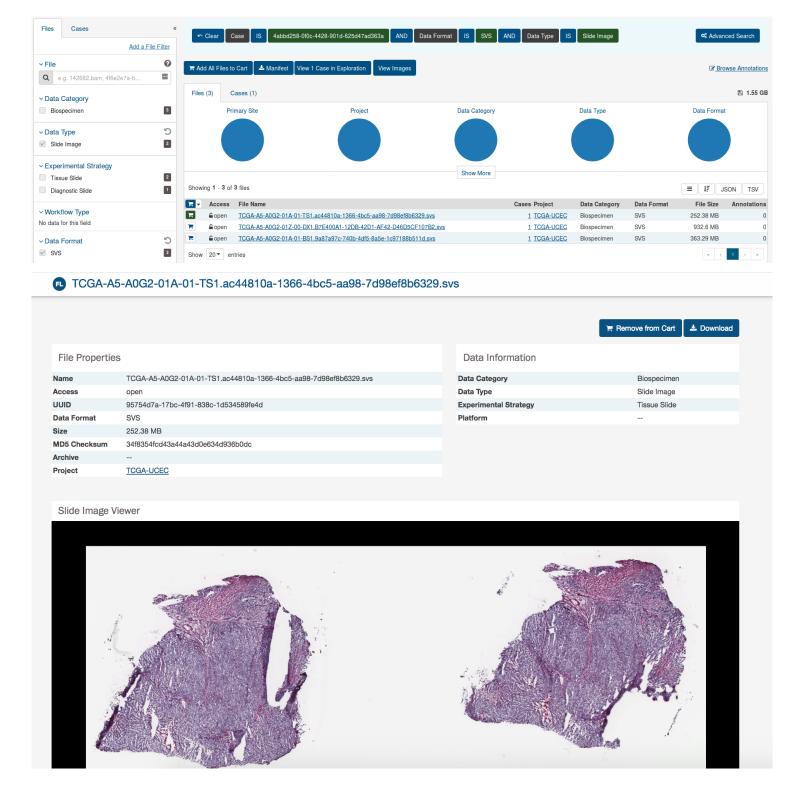
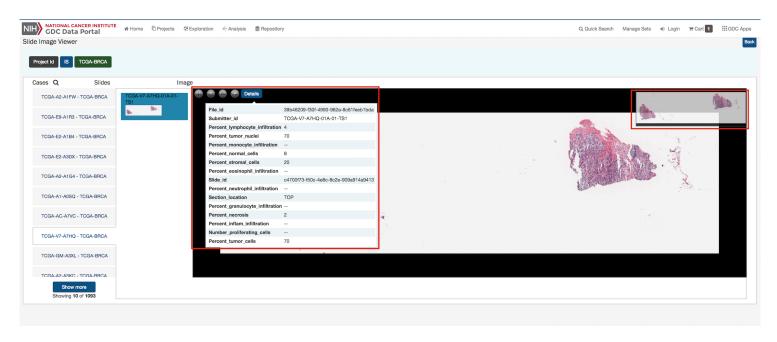


Image Viewer Features

In the image viewer, a user can:

- Zoom in and zoom out by clicking on + and icons.
- Reset to default display by clicking on the Home icon.
- Display the image in full screen mode by clicking on the Expand icon.
- View the slide detail by clicking on "Details" button.

• Selecting the area of interest with the thumbnail at the top-right corner.



Chapter 5

Genes and Mutations

Gene and Mutation Summary Pages

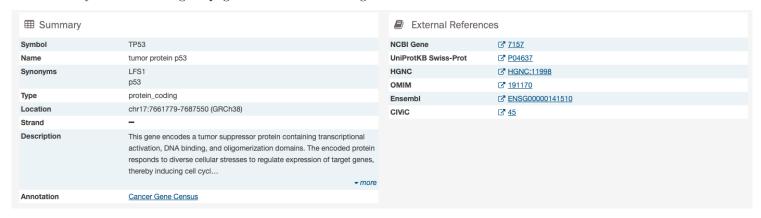
Many parts of the GDC website contain links to Gene and Mutation summary pages. These pages display information about specific genes and mutations, along with visualizations and data showcasing the relationship between themselves, the projects, and cases within the GDC. The gene and mutation data that is visualized on these pages are produced from the Open-Access MAF files available for download on the GDC Portal.

Gene Summary Page

Gene Summary Pages describe each gene with mutation data and provides results related to the analyses that are performed on these genes.

Summary

The summary section of the gene page contains the following information:



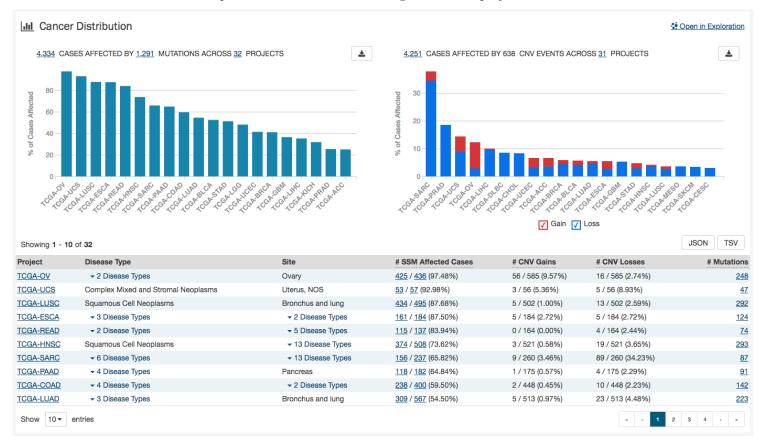
- Symbol: The gene symbol
- Name: Full name of the gene
- Synonyms: Synonyms of the gene name or symbol, if available
- Type: A broad classification of the gene
- Location: The chromosome on which the gene is located and its coordinates
- Strand: If the gene is located on the forward (+) or reverse (-) strand
- **Description:** A description of gene function and downstream consequences of gene alteration
- Annotation: A notation/link that states whether the gene is part of The Cancer Gene Census

External References

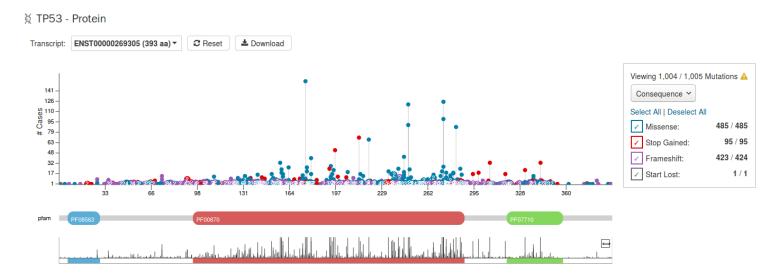
A list with links that lead to external databases with additional information about each gene is displayed here. These external databases include: Entrez, Uniprot, Hugo Gene Nomenclature Committee, Online Mendelian Inheritance in Man, Ensembl, and CIViC.

Cancer Distribution

A table and two bar graphs (one for mutations, one for CNV events) show how many cases are affected by mutations and CNV events within the gene as a ratio and percentage. Each row/bar represents the number of cases for each project. The final column in the table lists the number of unique mutations observed on the gene for each project.



Protein Viewer



Mutations and their frequency across cases are mapped to a graphical visualization of protein-coding regions with a lollipop plot. Pfam domains are highlighted along the x-axis to assign functionality to specific protein-coding regions. The bottom track represents a view of the full gene length. Different transcripts can be selected by using the drop-down menu above the plot.

The panel to the right of the plot allows the plot to be filtered by mutation consequences or impact. The plot will dynamically change as filters are applied. Mutation consequence and impact is denoted in the plot by color.

Note: The impact filter on this panel will not display the annotations for alternate transcripts.

The plot can be viewed at different zoom levels by clicking and dragging across the x-axis, clicking and dragging across the bottom track, or double clicking the pfam domain IDs. The Reset button can be used to bring the zoom level back to its original position. The plot can also be exported as a PNG image, SVG image or as JSON formatted text by choosing the Download button above the plot.

Most Frequent Mutations

The 20 most frequent mutations in the gene are displayed as a bar graph that indicates the number of cases that share each mutation.



A table is displayed below that lists information about each mutation including:

• DNA Change: The chromosome and starting coordinates of the mutation are displayed along with the nucleotide differences between the reference and tumor allele

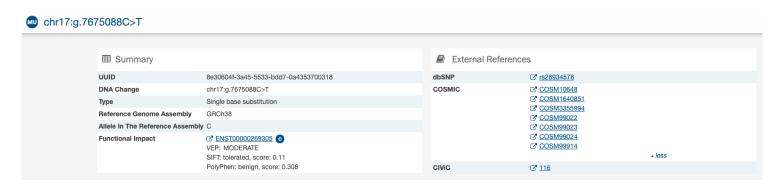
- Type: A general classification of the mutation
- Consequences: The effects the mutation has on the gene coding for a protein (i.e. synonymous, missense, non-coding transcript)
- ___ Affected Cases in Gene:___ The number of affected cases, expressed as number across all mutations within the Gene
- ___ Affected Cases Across GDC:___ The number of affected cases, expressed as number across all projects. Choosing the arrow next to the percentage will expand the selection with a breakdown of each affected project
- Impact: A subjective classification of the severity of the variant consequence. This determined using Ensembl VEP, PolyPhen, and SIFT. The categories are outlined here.

Note: The Mutation UUID can be displayed in this table by selecting it from the drop-down represented by three parallel lines Clicking the Open in Exploration button will navigate the user to the Exploration page, showing the same results in the table (mutations filtered by the gene).

Mutation Summary Page

The Mutation Summary Page contains information about one somatic mutation and how it affects the associated gene. Each mutation is identified by its chromosomal position and nucleotide-level change.

Summary



- **ID:** A unique identifier (UUID) for this mutation
- DNA Change: Denotes the chromosome number, position, and nucleotide change of the mutation
- Type: A broad categorization of the mutation
- Reference Genome Assembly: The reference genome in which the chromosomal position refers to
- Allele in the Reference Assembly: The nucleotide(s) that compose the site in the reference assembly
- Functional Impact: A subjective classification of the severity of the variant consequence.

External References

A separate panel contains links to databases that contain information about the specific mutation. These include dbSNP, COSMIC, and CIViC.

Consequences

The consequences of the mutation are displayed in a table. The set of consequence terms, defined by the Sequence Ontology.

Example 2 Consequences

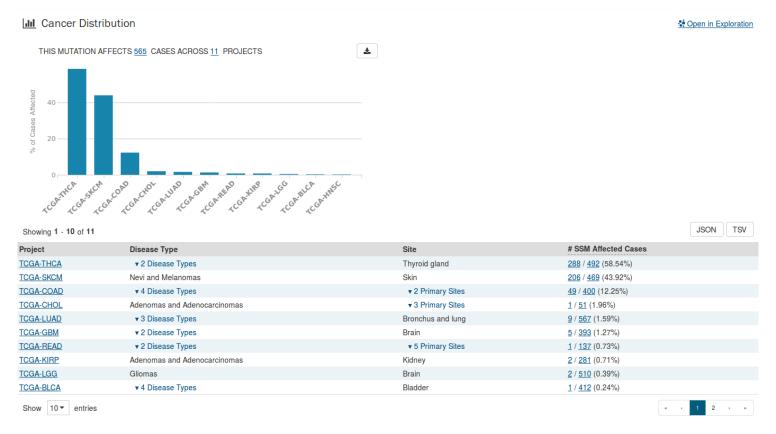


The fields that describe each consequence are listed below:

- **Gene:** The symbol for the affected gene
- AA Change: Details on the amino acid change, including compounds and position, if applicable
- Consequence: The biological consequence of each mutation
- Coding DNA Change: The specific nucleotide change and position of the mutation within the gene
- Strand: If the gene is located on the forward (+) or reverse (-) strand
- Transcript(s): The transcript(s) affected by the mutation. Each contains a link to the Ensembl entry for the transcript

Cancer Distribution

A table and bar graph shows how many cases are affected by the particular mutation. Each row/bar represents the number of cases for each project.



The table contains the following fields:

- Project ID: The ID for a specific project
- Disease Type: The disease associated with the project
- Site: The anatomical site affected by the disease
- ___ Affected Cases___: The number of affected cases and total number of cases displayed as a fraction and percentage

Protein Viewer



The protein viewer displays a plot representing the position of mutations along the polypeptide chain. The y-axis represents the number of cases that exhibit each mutation, whereas the x-axis represents the polypeptide chain sequence. Pfam domains that were identified along the polypeptide chain are identified with colored rectangles labeled with pfam IDs. See the Gene Summary Page for additional details about the protein viewer.

The panel to the right of the plot allows the plot to be filtered by mutation consequences or impact. The plot will dynamically change as filters are applied. Mutation consequence and impact is denoted in the plot by color.

Note: The impact filter on this panel will not display the annotations for alternate transcripts.

The plot can be viewed at different zoom levels by clicking and dragging across the x-axis, clicking and dragging across the bottom track, or double clicking the pfam domain IDs. The Reset button can be used to bring the zoom level back to its original position. The plot can also be exported as a PNG image, SVG image or as JSON formatted text by choosing the Download button above the plot.

Chapter 6

Custom Set Analysis

Analysis

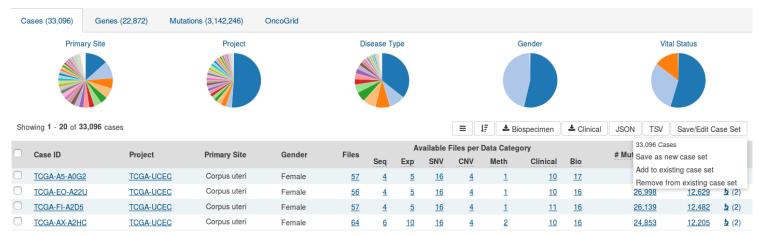
In addition to the Exploration Page, the GDC Data Portal also has features used to save and compare sets of cases, genes, and mutations. These sets can either be generated with existing filters (e.g. males with lung cancer) or through custom selection (e.g. a user-generated list of case IDs).

Note that saving a set only saves the type of entity included in the set. For example, a saved case set will not include filters that were applied to genes or mutations. Please be aware that your custom sets are deleted during each new GDC data release. You can export them and re-upload them in the "Manage Sets" link at the top right of the Portal.

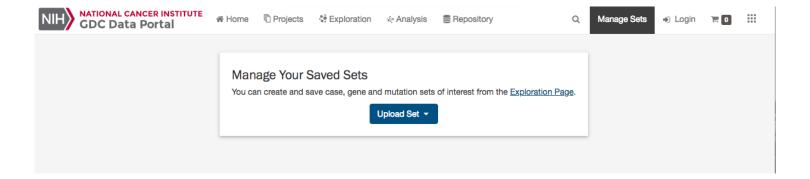
Generating a Cohort for Analysis

Cohort sets are completely customizable and can be generated for cases, genes, or mutations using the following methods:

Apply Filters in Exploration: Sets can be assembled using the existing filters in the Exploration page. They can be saved by choosing the "Save/Edit Case Set" button under the pie charts for case sets. This will prompt a decision to save as new case set. The same can be done for both gene and mutation filters, and can be applied and saved in the Genes and Mutations tab, respectively.

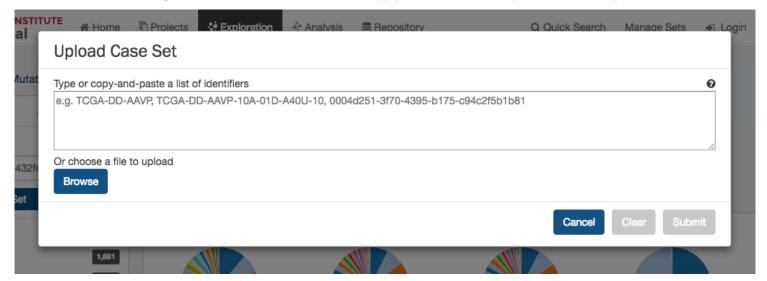


Upload ID Set: This feature is available in the "Manage Sets" link at the top right of the Portal. Choose "Upload Set" and then select whether the set comprises cases, genes, or mutations. A set of IDs or UUIDs can then be uploaded in a text file or copied and pasted into the list of identifiers field along with a name identifying the set. Once the list of identifiers is uploaded, the IDs are validated and grouped according to whether or not the identifier matched an existing GDC ID.

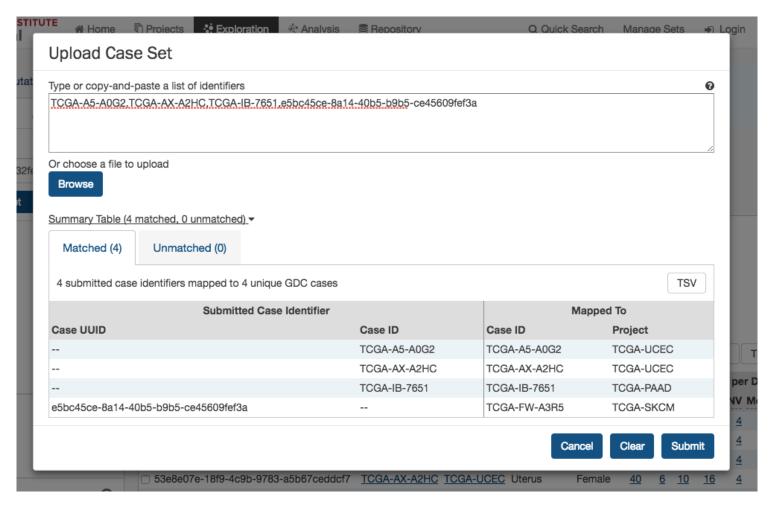


Upload Case Set

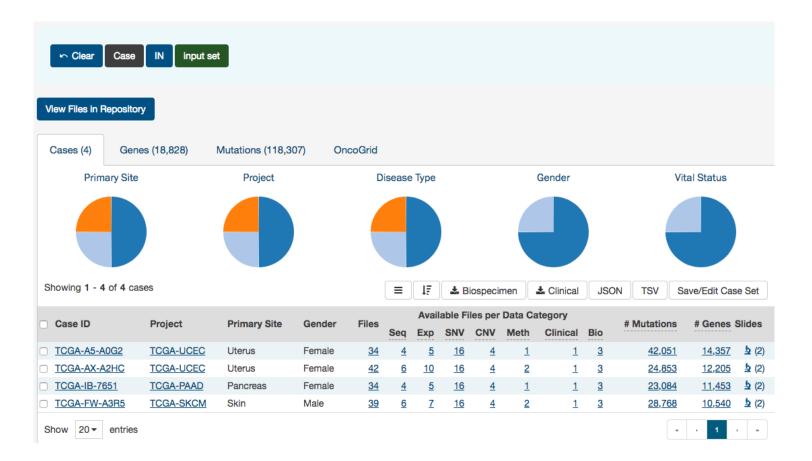
In the Cases filters panel, instead of supplying cases one-by-one, users can supply a list of cases. Clicking on the Upload Case Set button will launch a dialog as shown below, where users can supply a list of cases or upload a comma-separated text file of cases.



After supplying a list of cases, a table below will appear which indicates whether the case was found.

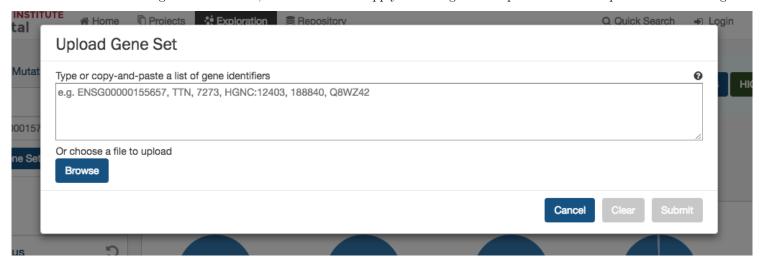


Clicking on Submit will filter the results in the Exploration Page by those cases.

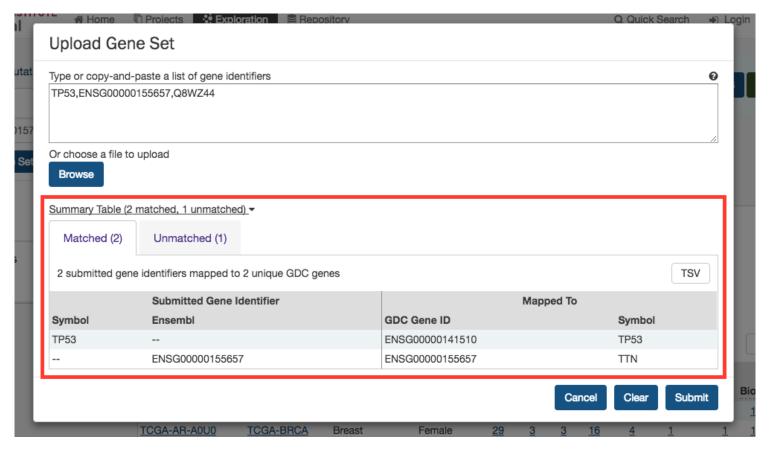


Upload Gene Set

In the Genes filters panel, instead of supplying genes one-by-one, users can supply a list of genes. Clicking on the Upload Gene Set button will launch a dialog as shown below, where users can supply a list of genes or upload a comma-separated text file of genes.



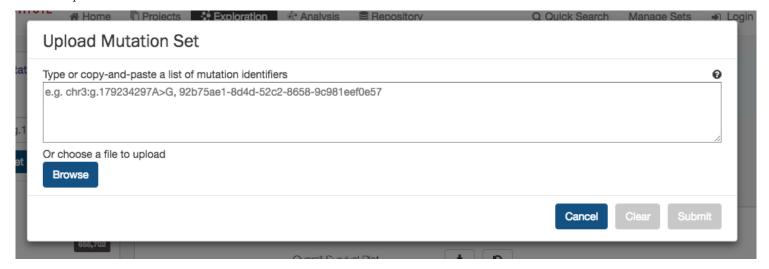
After supplying a list of genes, a table below will appear which indicates whether the gene was found.



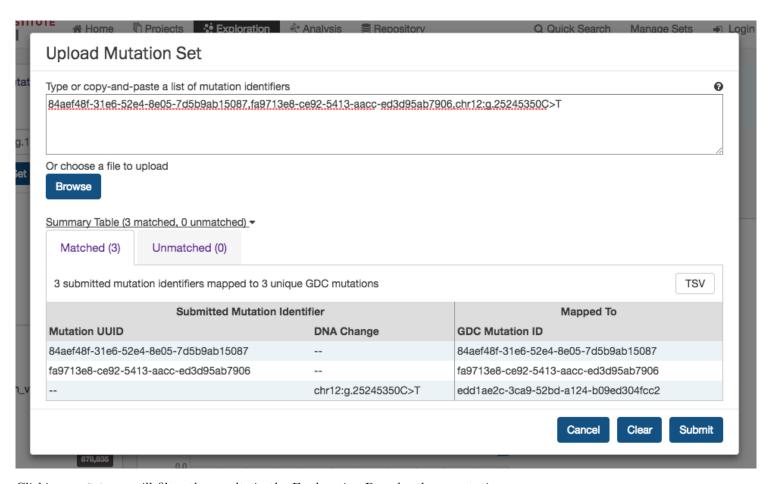
Clicking on Submit will filter the results in the Exploration Page by those genes.

Upload Mutation Set

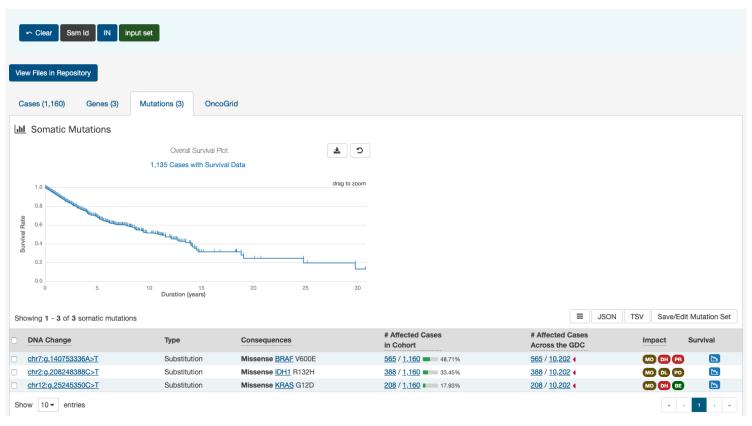
In the Mutations filters panel, instead of supplying mutation id's one-by-one, users can supply a list of mutations. Clicking on the Upload Mutation Set button will launch a dialog as shown below, where users can supply a list of mutations or upload a comma-separated text file of mutations.



After supplying a list of mutations, a table below will appear which indicates whether the mutation was found.

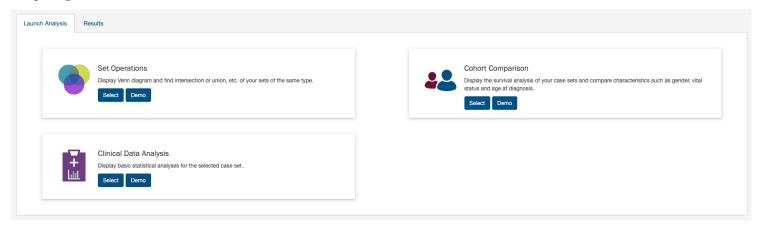


Clicking on Submit will filter the results in the Exploration Page by those mutations.



Analysis Page

Clicking on the Analysis button in the top toolbar will launch the Analysis Page which displays the various options available for comparing saved sets.

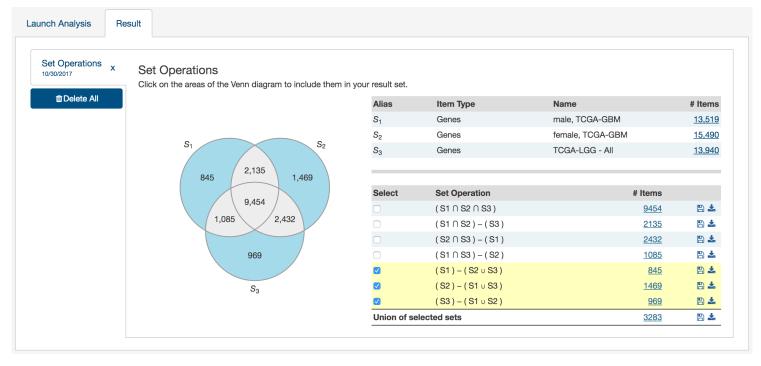


There are three tabs on this page:

- Launch Analysis: Where users can select either to do Set Operations, Cohort Comparison or Clinical Data Analysis.
- Results: Where users can view the results of current or previous set analyses.

Analysis Page: Set Operations

Up to three sets of the same set type can be compared and exported based on complex overlapping subsets. The features of this page include:



• Venn Diagram: Visually displays the overlapping items included within the three sets. Subsets based on overlap can be selected by clicking one or many sections of the Venn diagram. As sections of the Venn Diagram become highlighted in blue, their corresponding row in the overlap table becomes highlighted.

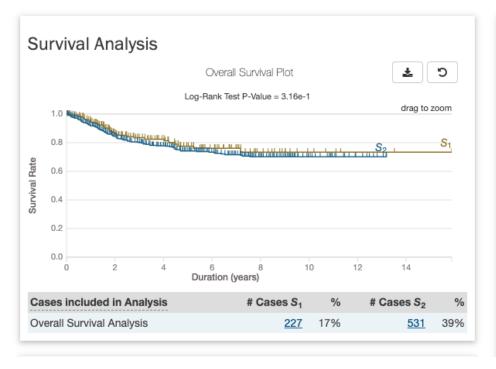
- Summary Table: Displays the alias, item type, and name for each set included in this analysis.
- Overlap Table: Displays the number of overlapping items with set operations rather than a visual diagram. Subsets can be selected by checking boxes in the "Select" column, which will highlight the corresponding section of the Venn Diagram. As rows are selected, the "Union of selected sets" row is populated. Each row has an option to save the subset as a new set, export the set as a TSV, or view files in the repository. The links that correspond to the number of items in each row will open the cohort in the Exploration Page.

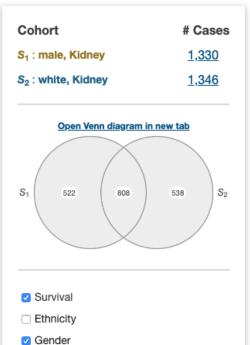
Analysis Tab: Cohort Comparison

The "Cohort Comparison" analysis displays a series of graphs and tables that demonstrate the similarities and differences between two case sets. The following features are displayed for each two sets:

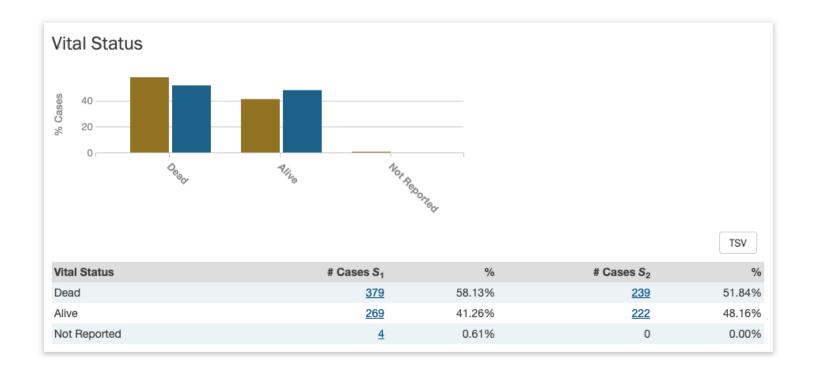
- A key detailing the number of cases in each cohort and the color that represents each (blue/gold).
- A Venn diagram, which shows the overlap between the two cohorts. The Venn diagram can be opened in a 'Set Operations' tab by choosing "Open Venn diagram in new tab".
- A selectable survival plot that compares both sets with information about the percentage of represented cases.

Cohort Comparison





• A breakdown of each cohort by selectable clinical facets with a bar graph and table. Facets include vital_status, gender, race, ethnicity, and age_at_diagnosis. A p-value (if it can be calculated from the data) that demonstrates whether the statuses are proportionally represented is displayed for the vital_status, gender, and ethnicity facets.



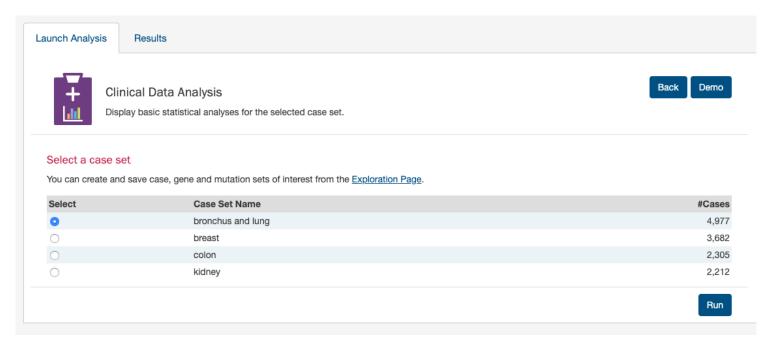
Analysis Tab: Clinical Data Analysis

The "Clinical Data Analysis" feature allows users to specifically examine the clinical data of a single case set in more detail. Users can select which clinical fields they want to display and visualize the data using various supported plot types. The clinical analysis features include:

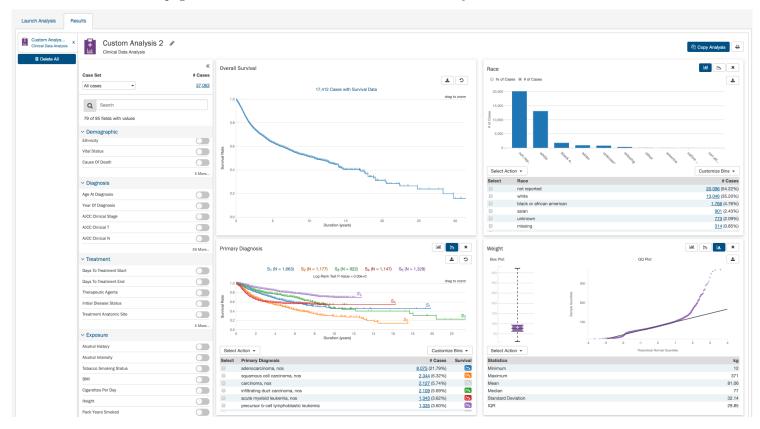
- Ability to select which clinical fields to display
- Examine the clinical data of each field using these visualizations:
 - Histogram
 - Survival Plot
 - Box Plot
 - QQ Plot
- Create custom bins for each field and re-visualize the data with those bins
- Select specific cases from a clinical field and use them to create a new set, or modify/remove from an existing set
- Download the visualizations of each plot type for each variable in SVG, PNG, JSON formats
- Download the data table of each field in TSV format
- Print all clinical variable cards in the analysis with their active plot to a single PDF

Selecting a Case Set

First, a case set must be selected to run the clinical analysis on:

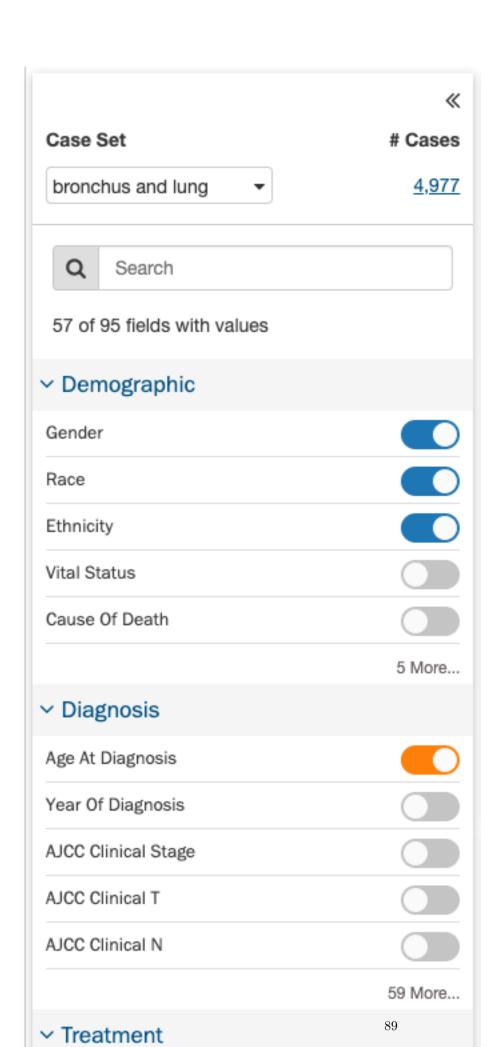


Click 'Run' - The results page loads with a new tab for the new clinical analysis for the selected set:



Enabling Clinical Variable Cards

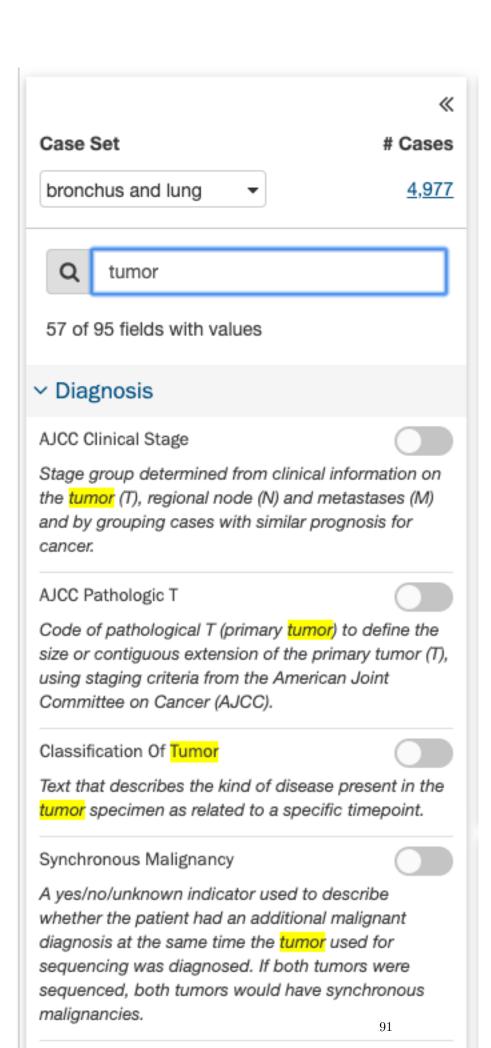
Users can use the control panel on the left side of the analysis to display which clinical variables they want. To enable or disable specific variables for display, click the on/off toggle controls:



The clinical fields are grouped into these categories:

- **Demographic:** Data for the characterization of the patient by means of segmenting the population (e.g. characterization by age, sex, race, etc.).
- **Diagnoses:** Data from the investigation, analysis, and recognition of the presence and nature of disease, condition, or injury from expressed signs and symptoms; also, the scientific determination of any kind; the concise results of such an investigation.
- **Treatments:** Records of the administration and intention of therapeutic agents provided to a patient to alter the course of a pathologic process.
- Exposures: Clinically-relevant patient information not immediately resulting from genetic predispositions.

Since the list of fields can be long, users can collapse and expand the field list for each clinical category for easier browsing, or use the search box:



Exploring Clinical Card Visualizations

Users can explore different visualizations for each clinical field they have enabled for display. Each card supports theses plot types:

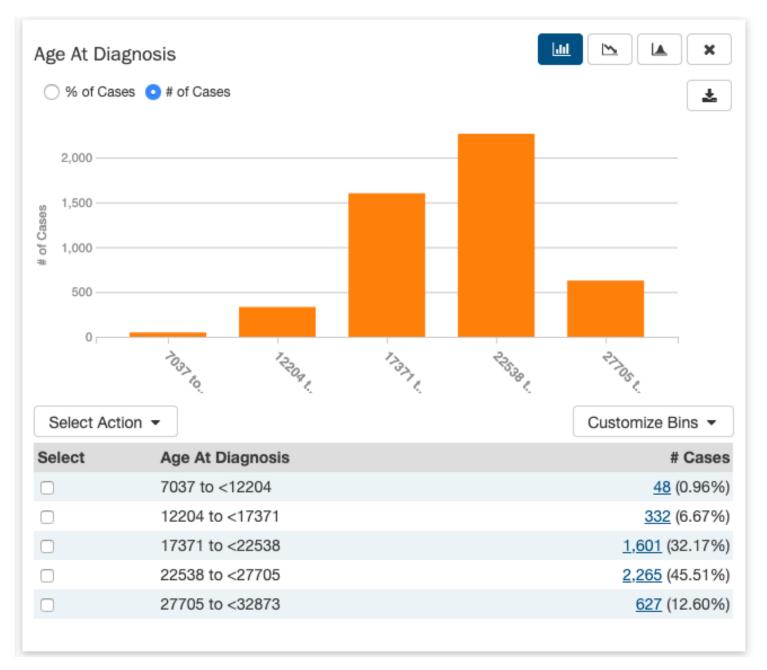
- Histogram
- Survival Plot
- Box Plot & QQ Plot (these plots are visualized side-by-side)

To switch between plot types, click the different plot type icons in the top-right of each card.

Histogram

The histogram plot type supports these features:

- View the distribution of cases (and % of cases) in the cohort for the clinical field's data categories as a histogram
- View the distribution of cases in tabular format
- Select the cases for specific data categories to create new sets, append to existing sets, or remove from existing sets
- Download the histogram visualization in SVG or PNG format
- Download the raw data used to generate the histogram in JSON format



Note that the histogram plot applies to, and can be displayed for, both categorical and continuous variables.

Survival Plot

The survival plot type supports these features:

- View the distribution of cases (and % of cases) in the cohort for the clinical field's data categories as a table
- Select and plot the survival analysis for the cases of specific data categories in the table:
 - By default the top 2 categories (highest of cases) are displayed
 - Users can manually select and plot up to 5 categories at a time
- Download the survival plot visualization in SVG or PNG format
- Download the raw data used to generate the survival plot in JSON or TSV format



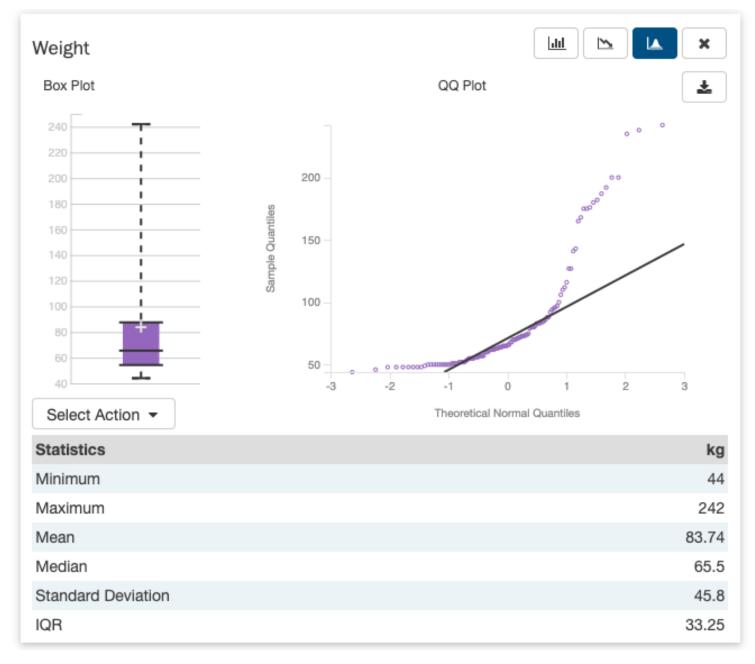
Note that the survival plot applies to, and can be displayed for, both categorical and continuous variables.

Box Plot & QQ Plot

The box plot and QQ plot are displayed side-by-side in the same visualization. This visualization supports these features:

- View standard summary statistics for the clinical field's data across the cohort as both a box plot visualization and a data table:
 - Minimum
 - Maximum
 - Mean
 - Median
 - Standard Deviation
 - Interquartile Range (IQR)
- View a QQ plot visualization to explore whether the clinical field's data across the cohort is normally distributed, where:

- The clinical data values are plotted as the sample quantiles on the vertical axis
- The the quantiles of the normal distribution are plotted on the horizontal axis
- Download the box plot and QQ plot visualizations in SVG or PNG format
- Download the raw data used to generate the QQ plot in JSON or TSV format



Note that the box plot and QQ plot only apply to continuous variables. They cannot be displayed for categorical variables.

Creating Custom Bins

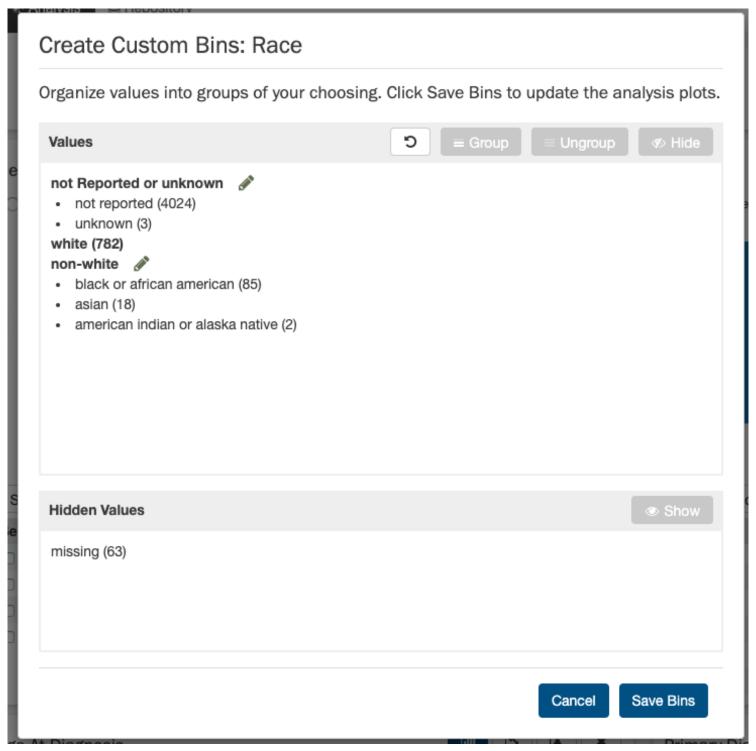
For each clinical variable, whether categorical or continuous, users can create custom bins to group the data in ways they find scientifically interesting or significant. Once saved, the bins are applied to these visualizations and they are then re-rendered:

- Histogram and associated data table
- Survival plot and associated data table

Custom bins can be reset to their defaults at any time for each card. Note that custom bins are saved per analysis.

Categorical Binning

To create custom bins for a categorical variable, click *Customize Bins*, then *Edit Bins*. A configuration window appears where the user can create their bins:



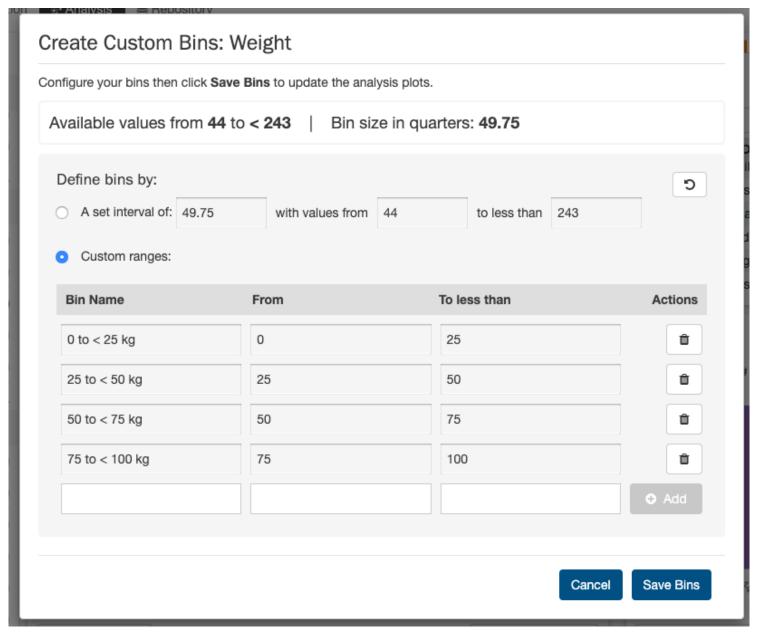
The user can:

- Group existing individual values into a single group
- Give a custom name to each group
- Ungroup previously grouped values
- Completely hide values from being shown in the visualization

• Re-show previously hidden values

Continuous Binning

To create custom bins for a continuous variable, click *Customize Bins*, then *Edit Bins*. A configuration window appears where the user can create their bins:

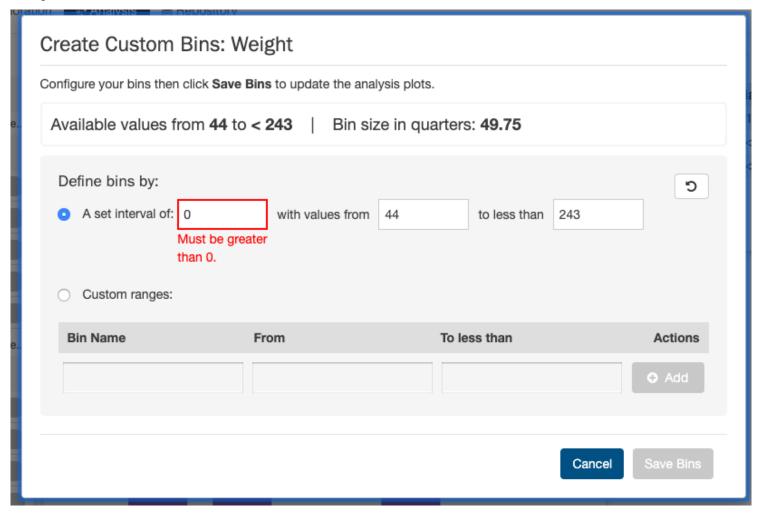


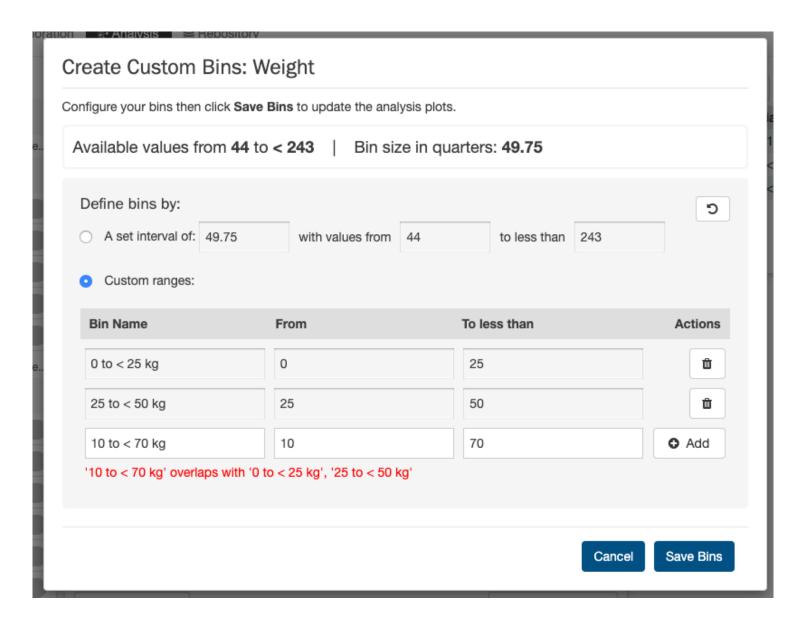
The user can choose one of these continuous binning methods:

- (1) Create equi-distant bins based on a set interval:
 - User must choose the interval (e.g. equi-distant bins of 1,825 days for the Age of Diagnosis field)
 - User can optionally define the starting and ending value between which the equi-distant bins will be created
- (2) Create completely custom ranges:
 - User manually enters 1 or more bins with custom ranges
 - User must enter a name for each range and the start and end values

- The ranges can be of different interval lengths

Before saving the bins, if there are errors in the configuration, the user will be notified to correct them and try saving again. For example:





Other Useful Functions

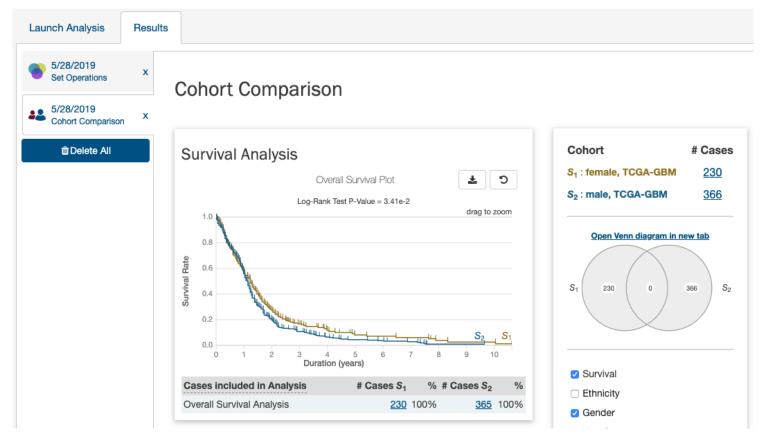
Clinical Analysis also provides these additional useful functions:

- Like other analysis types, all Clinical Analysis tabs are saved to the browser's local storage:
 - Each Analysis tab and its associated configurations (active cards, active plots, custom bins) is saved and is not deleted until local storage is cleared
 - The currently-enabled clinical cards and their currently-selected plot types are saved per analysis
 - Custom bins are saved per analysis
- Switch the current set that the analysis applies to This does the following:
 - Applies all currently-enabled clinical cards and their currently-selected plot types to the data in the set being switched to
 - Re-renders all active visualizations to reflect the data in the set being switched to
- Rename your analysis with a custom name
- Copy your current analysis to a new analysis:
 - User is prompted to name the new copy
 - All currently-enabled clinical cards and their currently-selected plot types are copied to the new analysis

- A new vertical tab appears on the left for the new copy
- Print the current analysis to PDF format:
 - All currently-enabled clinical cards and their currently-selected plot types are printed
 - 3 cards per page, with the fixed Overall Survival Plot displayed as the first card in the entire file

Analysis Page: Results

The results of the previous analyses are displayed on this page.



Each tab at the left side of the page is labeled according to the analysis type and the date that the analysis was performed and can be reviewed as long as it is present. The "Delete All" button will remove all of the previous analyses.

Chapter 7

Annotations

Annotations

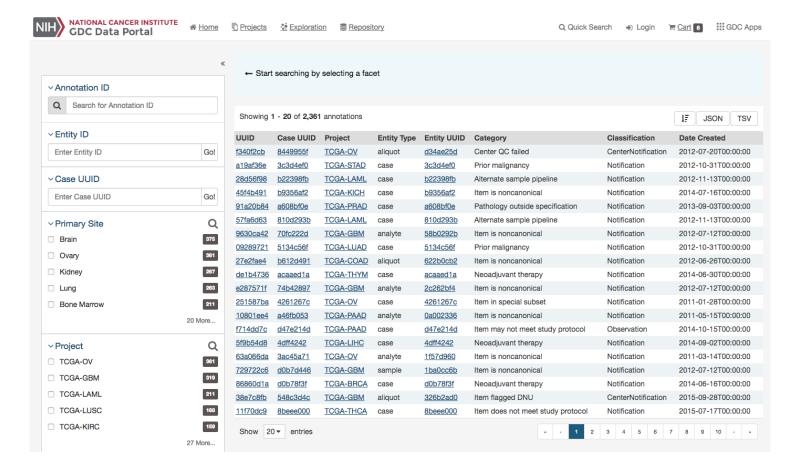
Annotations are notes added to individual cases, samples or files.

Annotations View

The Annotations View provides an overview of the available annotations and allows users to browse and filter the annotations based on a number of annotation properties (facets), such as the type of entity the annotation is attached to or the annotation category.

The view presents a list of annotations in tabular format on the right, and a facet panel on the left that allows users to filter the annotations displayed in the table. If facet filters are applied, the tabs on the right will display only the matching annotations. If no filters are applied, the tabs on the right will display information about all available data.

Clicking on an annotation ID in the annotations list will take the user to the Annotation Detail Page.



Facets Panel

The following facets are available to search for annotations:

- Annotation ID: Seach using annotation ID
- Entity ID: Seach using entity ID
- Case UUID: Seach using case UUID
- Primary Site: Anatomical site of the cancer
- Project: A cancer research project, typically part of a larger cancer research program
- Entity Type: The type of entity the annotation is associated with: Patient, Sample, Portion, Slide, Analyte, Aliquot
- Annotation Category: Search by annotation category.
- Annotation Created: Search for annotations by date of creation.
- Annotation Classification: Search by annotation classification.

Annotation Categories and Classification

For more details about categories and classifications please refer to the TCGA Annotations page on NCI Wiki.

Annotation Detail Page

The annotation entity page provides more details about a specific annotation. It is available by clicking on an annotation ID in Annotations View.





Q 🛊 Login 📜 🛭 🚻







AN f340f2cb-3cdc-5843-83d1-851d95d00f93

⊞ Summary	
Annotation UUID	f340f2cb-3cdc-5843-83d1-851d95d00f93
Entity UUID	d34ae25d-8073-4129-a1b8-3f43e19af73b
Entity Barcode	TCGA-24-1923-01A-01R-1567-13
Entity Type	aliquot
Case UUID	8449955f-42d9-46f6-919a-5cc5d59e6284
Case Submitter ID	TCGA-24-1923
Project ID	TCGA-OV
Classification	CenterNotification
Category	Center QC failed
Created On	2012-07-20T00:00:00
Status	Approved

NOTES

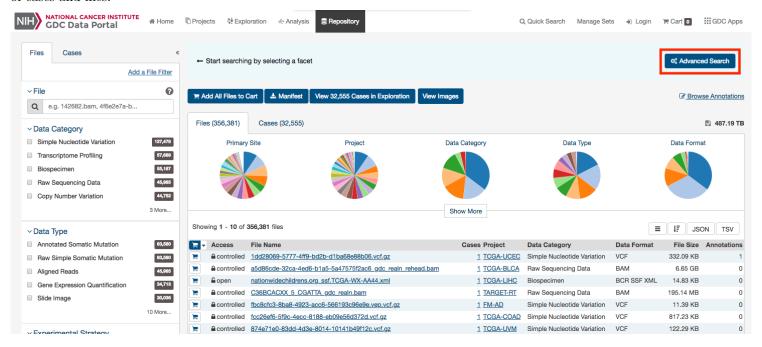
RNA-seq:LOW 5/3 COVERAGE RATIO

Chapter 8

Advanced Search

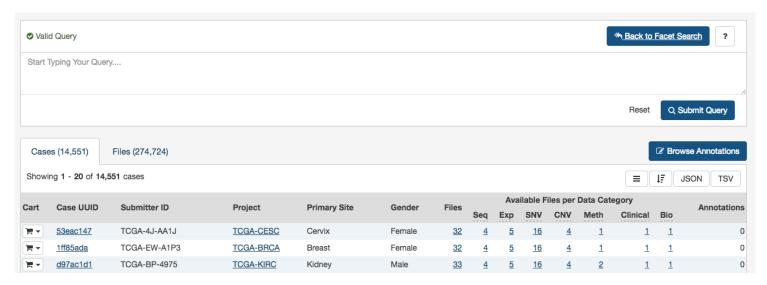
Advanced Search

Only available in the Repository view, the Advanced Search page offers complex query building capabilities to identify specific set of cases and files.

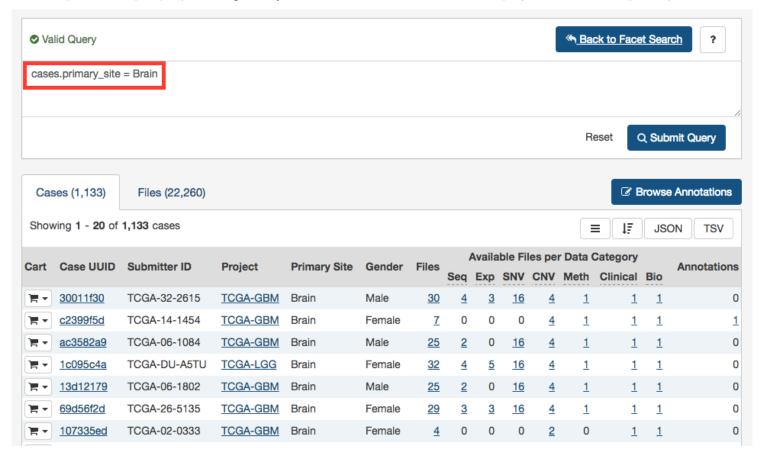


Overview: GQL

Advanced search allows for structured queries to search for files and cases. This is done via Genomic Query Language (GQL), a query language created by the GDC and OICR.



A simple query in GQL (also known as a 'clause') consists of a **field**, followed by an **operator**, followed by one or more **values**. For example, the simple query **cases.primary_site** = **Brain** will find all cases for projects in which the primary site is Brain:



Note that it is not possible to compare two fields (e.g. disease_type = project.name).

Note: GQL is not a database query language. For example, GQL does not have a "SELECT" statement.

Switching between Advanced Search and Facet Filters

When accessing Advanced Search from Repository View, a query created using facet filters in Repository View will be automatically translated to an Advanced Search GQL Query.

A query created in Advanced Search is not translated back to facet filters. Clicking on "Back to Facet Search" will return the user to Data View and reset the filters.

Using the Advanced Search

When opening the Advanced Search Page (via the Repository View), the search field will be automatically populated with facets filters already applied (if any).

This default query can be removed by pressing "Reset".

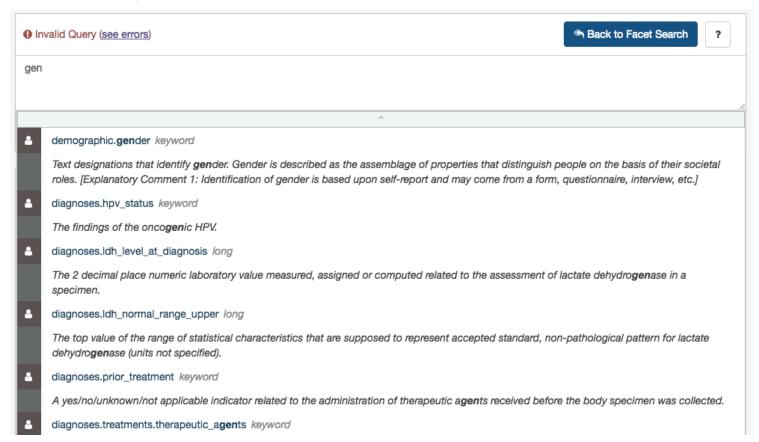
Once the query has been entered and is identified as a "Valid Query", click on "Search" to run your query.

Auto-complete

As a query is being written, the GDC Data Portal will analyze the context and offer a list of auto-complete suggestions. Auto-complete suggests both fields and values as described below.

Field Auto-complete

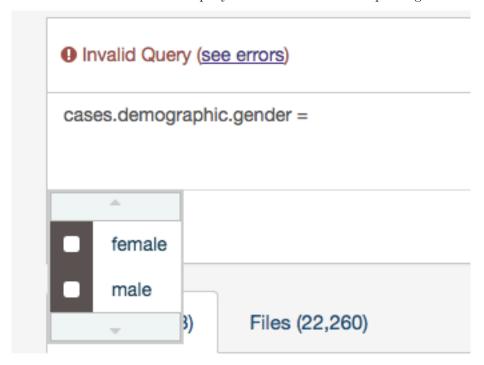
The list of auto-complete suggestions includes **all** available fields matching the user text input. The user has to scroll down to see more fields in the dropdown:



Value Auto-complete

The list of auto-complete suggestions includes top 100 values that match the user text input. The user has to scroll down to see more values in the dropdown.

The value auto-complete is not aware of the general context of the query, the system will display all available values in GDC for the selected field. It means the query could return 0 results depending of other filters.



Note: Quotes are automatically added to the value if it contains spaces.

Setting Precedence of Operators

You can use parentheses in complex GQL statements to enforce the precedence of operators.

For example, if you want to find all the open files in TCGA program as well as the files in TARGET program, you can use parentheses to enforce the precedence of the Boolean operators in your query, i.e.:

1 (files.access = open and cases.project.program.name = TCGA) or cases.project.program.name = TARGET

Note: Without parentheses, the statement will be evaluated left-to-right.

Keywords

A GQL keyword is a word that joins two or more clauses together to form a complex GQL query.

List of Keywords:

- AND
- OR

Note: Parentheses can be used to control the order in which clauses are executed.

"AND" Keyword

Used to combine multiple clauses, allowing you to refine your search.

Examples:

- Find all open files in breast cancer:
- 1 cases.primary_site = Breast and files.access = open
- Find all open files in breast cancer and data type is gene expression quantification:
- 1 cases.primary_site = Breast and files.access = open and files.data_type = "Gene Expression Quantification"

"OR" Keyword

Used to combine multiple clauses, allowing you to expand your search.

Note: The IN keyword can be an alternative to OR and result in simplified queries.

Examples:

- Find all files that are raw sequencing data or aligned reads:
- 1 files.data_type = "Aligned Reads" or files.data_type = "Raw sequencing data"
- Find all files where cases are male or vital status is alive:
- 1 cases.demographic.gender = male or cases.diagnoses.vital_status = alive

Operators

An operator in GQL is one or more symbols or words comparing the value of a field on its left with one or more values on its right, such that only true results are retrieved by the clause.

List of Operators and Query format

Operator	Description	
=	Field EQUAL Value (String or Number)	
!=	Field NOT EQUAL Value (String or Number)	
<	Field LOWER THAN Value (Number or Date)	
<=	Field LOWER THAN OR EQUAL Value (Number or Date)	
>	Field GREATER THAN Value (Number or Date)	
>=	Field GREATER THAN OR EQUAL Value (Number or Date)	
N	Field IN [Value 1, Value 2]	
EXCLUDE	Field EXCLUDE [Value 1, Value 2]	
S MISSING	Field IS MISSING	
NOT MISSING	Field NOT MISSING	

"=" Operator - EQUAL

The "=" operator is used to search for files where the value of the specified field exactly matches the specified value.

Examples:

- Find all files that are gene expression quantification:
- 1 files.data_type = "Gene Expression Quantification"
- Find all cases whose gender is female:
- 1 cases.demographic.gender = female

"!=" Operator - NOT EQUAL

The "!=" operator is used to search for files where the value of the specified field does not match the specified value.

The "!=" operator will not match a field that has no value (i.e. a field that is empty). For example:

```
cases.demographic.gender != male
```

This search will only match cases who have a gender and the gender is not male. To find cases other than male or with no gender populated, you would need to search:

```
cases.demographic.gender != male or cases.demographic.gender is missing.
```

Example:

- Find all files with an experimental strategy that is not genotyping array:
- 1 files.experimental_strategy != "Genotyping array"

">" Operator - GREATER THAN

The ">" operator is used to search for files where the value of the specified field is greater than the specified value.

Example:

- Find all cases whose number of days to death is greater than 60:
- 1 cases.diagnoses.days_to_death > 60

">=" Operator - GREATER THAN OR EQUALS

The ">=" operator is used to search for files where the value of the specified field is greater than or equal to the specified value.

Example:

- Find all cases whose number of days to death is equal or greater than 60:
- 1 cases.diagnoses.days_to_death >= 60

"<" Operator - LESS THAN

The "<" operator is used to search for files where the value of the specified field is less than the specified value.

Example:

- Find all cases whose age at diagnosis is less than 400 days:
- 1 cases.diagnoses.age_at_diagnosis < 400</pre>

"<=" Operator - LESS THAN OR EQUALS

The "<=" operator is used to search for files where the value of the specified field is less than or equal to the specified value. Example:

• Find all cases with a number of days to death less than or equal to 20:

```
1 cases.diagnoses.days_to_death <= 20</pre>
```

"IN" Operator

The "IN" operator is used to search for files where the value of the specified field is one of multiple specified values. The values are specified as a comma-delimited list, surrounded by brackets [].

Using "IN" is equivalent to using multiple "=" (EQUALS) statements, but is shorter and more convenient. That is, these two following statement will retrieve the same output:

```
cases.project.name IN [ProjectA, ProjectB, ProjectC]
cases.project.name = "ProjectA" OR cases.project.name = "ProjectB" OR cases.project.name = "ProjectC"
```

Examples:

- Find all files in breast, brain, and lung cancer:
- 1 cases.primary_site IN [Breast, Brain, Lung]
- Find all files that are annotated somactic mutations or raw simple somatic mutations:
- 1 files.data_type IN ["Annotated Somatic Mutation", "Raw Simple Somatic Mutation"]

"EXCLUDE" Operator

The "EXCLUDE" operator is used to search for files where the value of the specified field is not one of multiple specified values.

Using "EXCLUDE" is equivalent to using multiple "!=" (NOT_EQUALS) statements, but is shorter and more convenient. That is, these two following statement will retrieve the same output:

```
cases.project.name EXCLUDE [ProjectA, ProjectB, ProjectC]
cases.project.name != "ProjectA" OR cases.project.name != "ProjectB" OR cases.project.name !=
"ProjectC"
```

The "EXCLUDE" operator will not match a field that has no value (i.e. a field that is empty). For example:

```
files.experimental_strategy EXCLUDE ["WGS","WXS"]
```

This search will only match files that have an experimental strategy **and** the experimental strategy is not "WGS" or "WXS". To find files with an experimental strategy different than "WGS" or "WXS" **or is not assigned**, you would need to type:

```
files.experimental_strategy in ["WXS","WGS"] or files.experimental_strategy is missing
```

Examples:

- Find all files where experimental strategy is not WXS, WGS, Genotyping array:
- 1 files.experimental_strategy EXCLUDE [WXS, WGS, "Genotyping array"]

"IS MISSING" Operator

The "IS" operator can only be used with "MISSING". That is, it is used to search for files where the specified field has no value. Examples:

- Find all cases where gender is missing:
- 1 cases.demographic.gender is MISSING

"NOT MISSING" Operator

The "NOT" operator can only be used with "MISSING". That is, it is used to search for files where the specified field has a value.

Examples:

- Find all cases where race is not missing:
- 1 cases.demographic.race NOT MISSING

Special Cases

Date Format

The date format should be the following: YYYY-MM-DD (without quotes).

Example:

files.updated_datetime > 2015-12-31

Using Quotes

A value must be quoted if it contains a space. Otherwise the advanced search will not be able to interpret the value.

Quotes are not necessary if the value consists of one single word.

- Example: Find all cases with primary site is brain and data type is copy number segment:
- 1 cases.primary_site = Brain and files.data_type = "Copy Number Segment"

Age at Diagnosis - Unit in Days

The unit for age at diagnosis is in days. The user has to convert the number of years to number of days.

The **conversion factor** is 1 year = 365.25 days

- Example: Find all cases whose age at diagnosis > 40 years old (40 * 365.25)
- 1 cases.diagnoses.age_at_diagnosis > 14610

Fields Reference

The full list of fields available on the GDC Data Portal can be found through the GDC API using the following endpoint:

https://api.gdc.cancer.gov/gql/_mapping

Chapter 9

Authentication

Authentication

Overview

The GDC Data Portal provides granular metadata for all datasets available in the GDC. Any user can see a listing of all available data files, including controlled-access files. The GDC Data Portal also allows users to download open-access files without logging in. However, downloading of controlled-access files is restricted to authorized users and requires authentication.

Logging into the GDC

To login to the GDC, users must click on the Login button on the top right of the GDC website.

After clicking Login, users authenticate themselves using their eRA Commons login and password. If authentication is successful, the eRA Commons username will be displayed in the upper right corner of the screen, in place of the "Login" button.

Upon successful authentication, GDC Data Portal users can:

- see which controlled-access files they have access to;
- download controlled-access files directly from the GDC Data Portal;
- download an authentication token for use with the GDC Data Transfer Tool or the GDC API.

Controlled-access files are identified using a "lock" icon:

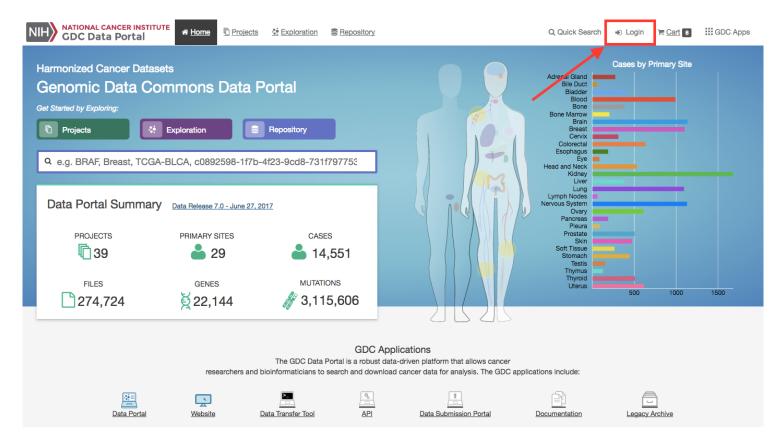
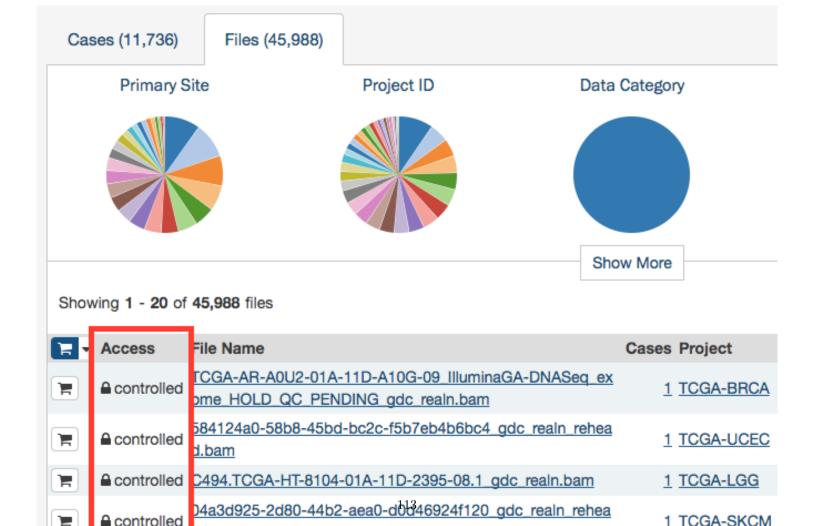


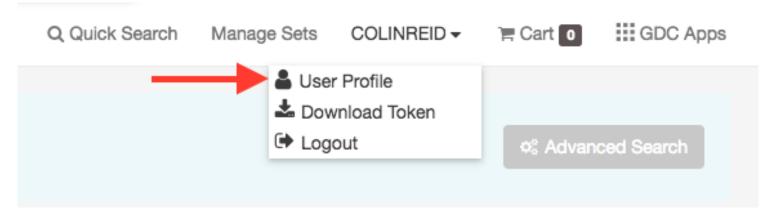
Figure 9.1: Login



The rest of this section describes controlled data access features of the GDC Data Portal available to authorized users. For more information about open and controlled-access data, and about obtaining access to controlled data, see Data Access Processes and Tools.

User Profile

After logging into the GDC Portal, users can view which projects they have access to by clicking the User Profile section in the dropdown menu in the top corner of the screen.



Clicking this button shows the list of projects.



GDC Authentication Tokens

The GDC Data Portal provides authentication tokens for use with the GDC Data Transfer Tool or the GDC API. To download a token:

- 1. Log into the GDC using your eRA Commons credentials
- 2. Click the username in the top right corner of the screen
- 3. Select the "Download token" option

A new token is generated each time the Download Token button is clicked.

For more information about authentication tokens, see Data Security.

NOTE: The authentication token should be kept in a secure location, as it allows access to all data accessible by the associated user account.

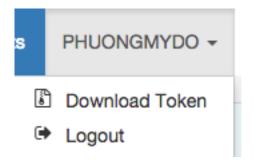


Figure 9.2: Token Download Button

Logging Out

To log out of the GDC, click the username in the top right corner of the screen, and select the Logout option.

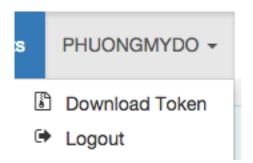


Figure 9.3: Logout link

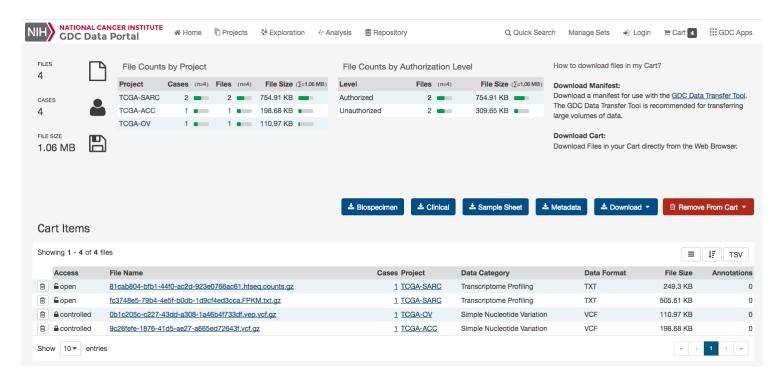
Chapter 10

File Cart

Cart and File Download

While browsing the GDC Data Portal, files can either be downloaded individually from File Summary Pages or collected in the file cart to be downloaded as a bundle. Clicking on the shopping cart icon that is next to any item in the GDC will add the item to your cart.

GDC Cart



Cart Summary

The Cart Summary Page shows a summary of all files currently in the cart:

- Number of files.
- Number of cases associated with the files.

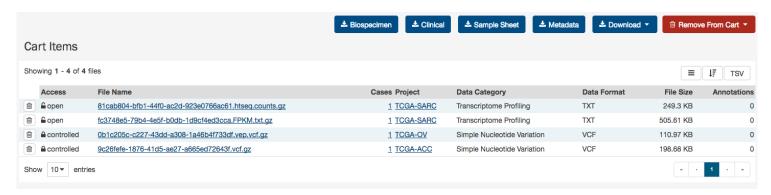
• Total file size.

The Cart page also displays two tables:

- File count by project: Breaks down the files and cases by each project.
- File count by authorization level: Breaks down the files in the cart by authorization level. A user must be logged into the GDC in order to download 'Controlled-Access files'.

The cart also directs users how to download files in the cart. For large data files, it is recommended that the GDC Data Transfer Tool be used.

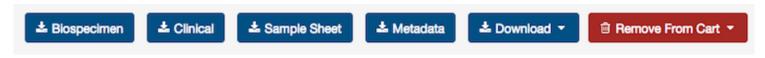
Cart Items



The Cart Items table shows the list of all the files that were added to the Cart. The table gives the following information for each file in the cart:

- Access: Displays whether the file is open or controlled access. Users must login to the GDC Portal and have the appropriate
 credentials to access these files.
- File Name: Name of the file. Clicking the link will bring the user to the File Summary Page.
- Cases: How many cases does the file contain. Clicking the link will bring the user to the Case Summary Page.
- Project: The Project that the file belongs to. Clicking the link will bring the user to the Project Summary Page.
- Category: Type of data.
- Format: The file format.
- Size: The size of the file.
- Annotations: Whether there are any annotations.

Download Options



The following buttons on the Cart page allows users to download files that are related to the ones in the cart. The following download options are available:

- Biospecimen: Downloads biospecimen data related to files in the cart in either TSV or JSON format.
- Clinical: Downloads clinical data related to files in the cart in either TSV or JSON format.
- Sample Sheet: Downloads a tab-separated file which contains the associated case/sample IDs and the sample type (Tumor/Normal) for each file in the cart.
- Metadata: GDC harmonized clinical, biospecimen, and file metadata associated with the files in the cart.
- Download:

- Manifest: Download a manifest file for use with the GDC Data Transfer Tool to download files. A manifest file contains a list of the UUIDs that correspond to the files in the cart.
- Cart: Download the files in the Cart directly through the browser. Users have to be cautious of the amount of data in the cart since this option will not optimize bandwidth and will not provide resume capabilities.
- Remove from Cart: Remove all files or unauthorized files from the cart.
- SRA XML, MAGE-TAB: This option is available in the GDC Legacy Archive only. It is used to download metadata files associated with the files in the cart.

The cart allows users to download up to 5 GB of data directly through the web browser. This is not recommended for downloading large volumes of data, in particular due to the absence of a retry/resume mechanism. For downloads over 5 GB we recommend using the Download Manifest button and download a manifest file that can be imported into GDC Data Transfer Tool.

Note: when downloading multiple files from the cart, they are automatically bundled into one single Gzipped (.tar.gz) file.

GDC Data Transfer Tool

The Download Manifest button will download a manifest file that can be imported into the GDC Data Transfer Tool. Below is an example of the contents of a manifest file used for download:

1	id filename md5 size state
2	4ea9c657-8f85-44d0-9a77-ad59cced8973 mdanderson.org_ESCA.MDA_RPPA_Core.mage-tab.1.1.0.tar.gz
	2516051 live
3	b8342cd5-330e-440b-b53a-1112341d87db mdanderson.org_SARC.MDA_RPPA_Core.mage-tab.1.1.0.tar.gz
	4523632 live
4	c57673ac-998a-4a50-a12b-4cac5dc3b72e mdanderson.org_KIRP.MDA_RPPA_Core.mage-tab.1.2.0.tar.gz
	4195746 live
5	3f22dd8d-59c8-43a4-89cf-3b595f2e5a06 14-3-3 beta-R-V GBL1112940.tif 56df0e4b4fc092fc3643bd2e316ac05b
	6257840 live
6	7ce05059-9197-4d38-830f-04356f5f851a 14-3-3 beta-R-V GBL11066140.tif 6abfee483974bc2e61a37b5499ae9a07
U	6261580 live
	0201000 1210
7	8e00d22a-ca6f-4da8-a1c3-f23144cb21b7 14-3-3_beta-R-V_GBL1112940.tif 56df0e4b4fc092fc3643bd2e316ac05b
	6257840 live
8	96487cd7-8fa8-4bee-9863-17004a70b2e9 14-3-3_beta-R-V_GBL1112940.tif 56df0e4b4fc092fc3643bd2e316ac05b
	6257840 live

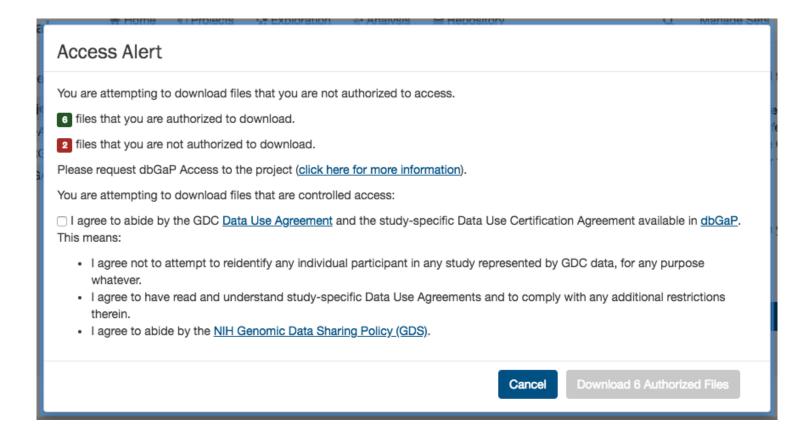
The Manifest contains a list of the file UUIDs in the cart and can be used together with the GDC Data Transfer Tool to download all files.

Information on the GDC Data Transfer Tool is available in the GDC Data Transfer Tool User's Guide.

Controlled Files

If a user tries to download a cart containing controlled files and without being authenticated, a pop-up will be displayed to offer the user either to download only open access files or to login into the GDC Data Portal through eRA Commons. See Authentication for details.

Once a user is logged in, controlled files that they have access to can be downloaded. To download files from the portal, users must agree to the GDC and individual project Data Use Agreements by selecting the agreement checkbox on the Access Alert message.



Authentication

The GDC Data Portal provides granular metadata for all datasets available in the GDC. Any user can see a listing of all available data files, including controlled-access files. The GDC Data Portal also allows users to download open-access files without logging in. However, downloading of controlled-access files is restricted to authorized users and requires authentication.

Logging into the GDC

To login to the GDC, users must click on the Login button on the top right of the GDC Website.

After clicking Login, users authenticate themselves using their eRA Commons login and password. If authentication is successful, the eRA Commons username will be displayed in the upper right corner of the screen, in place of the "Login" button.

Upon successful authentication, GDC Data Portal users can:

- See which controlled-access files they can access.
- Download controlled-access files directly from the GDC Data Portal.
- Download an authentication token for use with the GDC Data Transfer Tool or the GDC API.
- See controlled-access mutation data they can access.

Controlled-access files are identified using a "lock" icon:

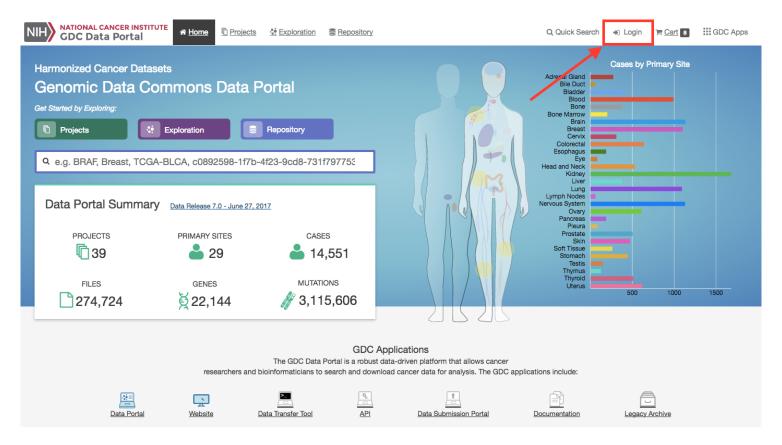
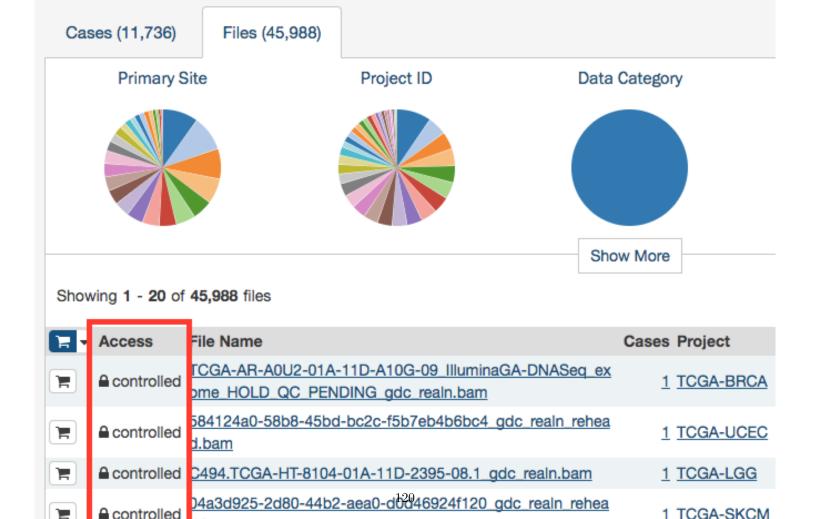


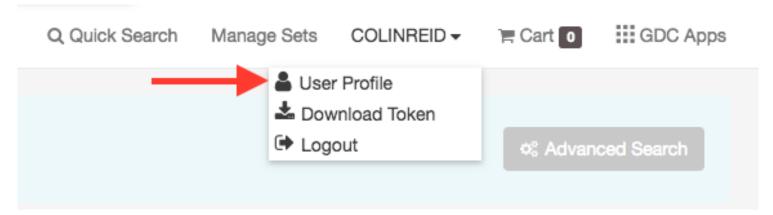
Figure 10.1: Login



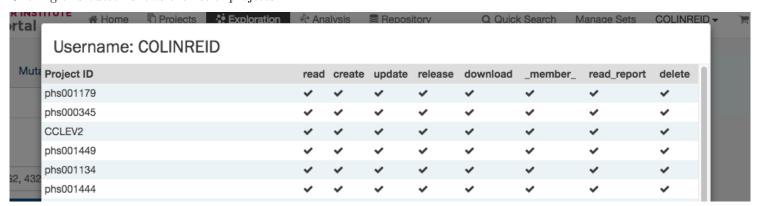
The rest of this section describes controlled data access features of the GDC Data Portal available to authorized users. For more information about open and controlled-access data, and about obtaining access to controlled data, see Data Access Processes and Tools.

User Profile

After logging into the GDC Portal, users can view which projects they have access to by clicking the User Profile section in the dropdown menu in the top corner of the screen.



Clicking this button shows the list of projects.



GDC Authentication Tokens

The GDC Data Portal provides authentication tokens for use with the GDC Data Transfer Tool or the GDC API. To download a token:

- 1. Log into the GDC using your eRA Commons credentials.
- 2. Click the username in the top right corner of the screen.
- 3. Select the "Download token" option.

A new token is generated each time the Download Token button is clicked.

For more information about authentication tokens, see Data Security.

Note: The authentication token should be kept in a secure location, as it allows access to all data accessible by the associated user account.

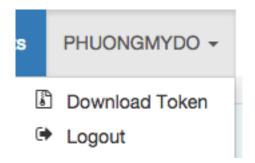


Figure 10.2: Token Download Button

Logging Out

To log out of the GDC, click the username in the top right corner of the screen, and select the Logout option.

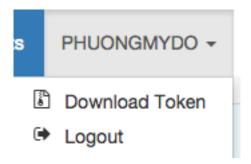


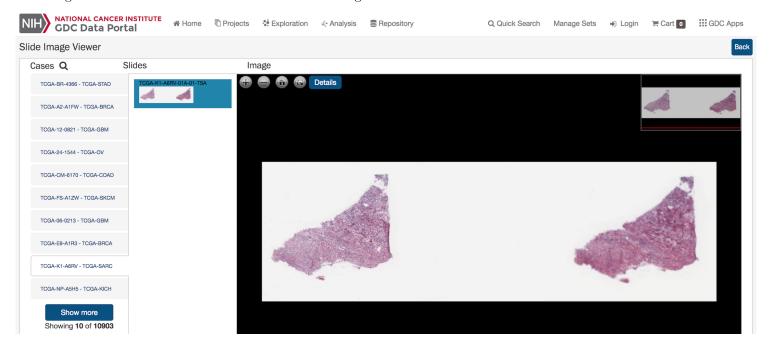
Figure 10.3: Logout link

Chapter 11

Image Viewer

Image Viewer

The Image viewer allows users to visualize tissue slide images.



How to access the image viewer

The image viewer is available from: * Case entity page: Click on the image viewer icon in the Case summary section or in the Biospecimen section - Slides detail. It will display in the image viewer the tissue slide images attached to the Case.

TCGA-UCEC / TCGA-A5-A0G2 Add all files to the cart **FILES III** Summary <u>34</u> **Case UUID** 4abbd258-0f0c-4428-901d-625d47ad363a Case ID TCGA-A5-A0G2 **ANNOTATIONS Project TCGA-UCEC** 0 **Project Name** Uterine Corpus Endometrial Carcinoma **Disease Type** Uterine Corpus Endometrial Carcinoma **Program TCGA Primary Site** Uterus **Images ½** (2) 📜 CA TCGA-UCEC / TCGA-A5-A0G2 Biospecimen ▲ Download Slide ID TCGA-A5-A0G2-01A-01-TS1 Q Search Collapse All Slide UUID ac44810a-1366-4bc5-aa98-7d98ef8b6329 Samples 20 **Percent Lymphocyte Infiltration △TCGA-A5-A0G2-10A Percent Tumor Nuclei** 90 ■ Portions **Percent Monocyte Infiltration** ∆TCGA-A5-A0G2-01A **Percent Normal Cells** 10 Portions ∆TCGA-A5-A0G2-01A-21-A162-20 **Percent Stromal Cells Percent Eosinophil Infiltration △TCGA-A5-A0G2-01A-11**

• Case table in Exploration and Repository pages: Click on the image viewer icon in the Case table. It will display in the image viewer all the tissue slide images attached to the Case.

Percent Neutrophil Infiltration

Percent Granulocyte Infiltration

Percent Inflam Infiltration
Number Proliferating Cells
Percent Tumor Cells

Section Location

Percent Necrosis

Slide Image

♣ Analytes➡ Slides

△TCGA-A5-A0G2-01A-01-TS1▶

△TCGA-A5-A0G2-01A-01-BS1

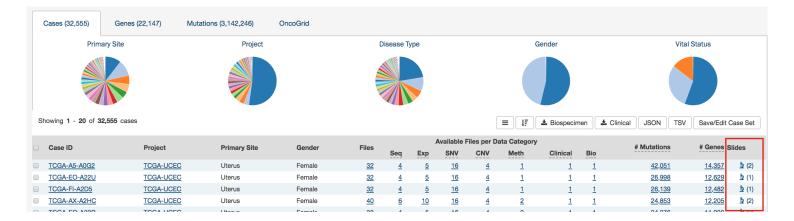
5

5

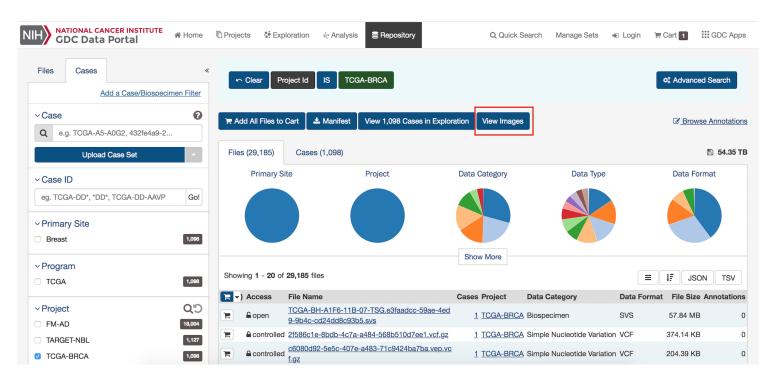
85

P 🗯 🖛

TOP



• Repository page - main search: After selecting your query, click on "View images" in Repository. It will display the tissue slide images of all the cases resulting from the query.

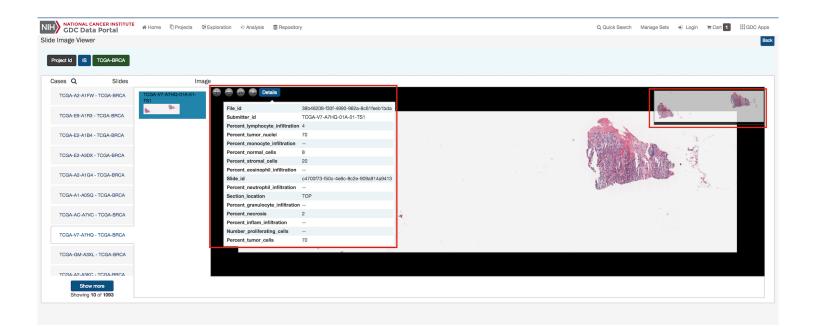


• Directly in the File entity page: You can visualize the tissue slide image directly in the file entity page.



Image viewer features

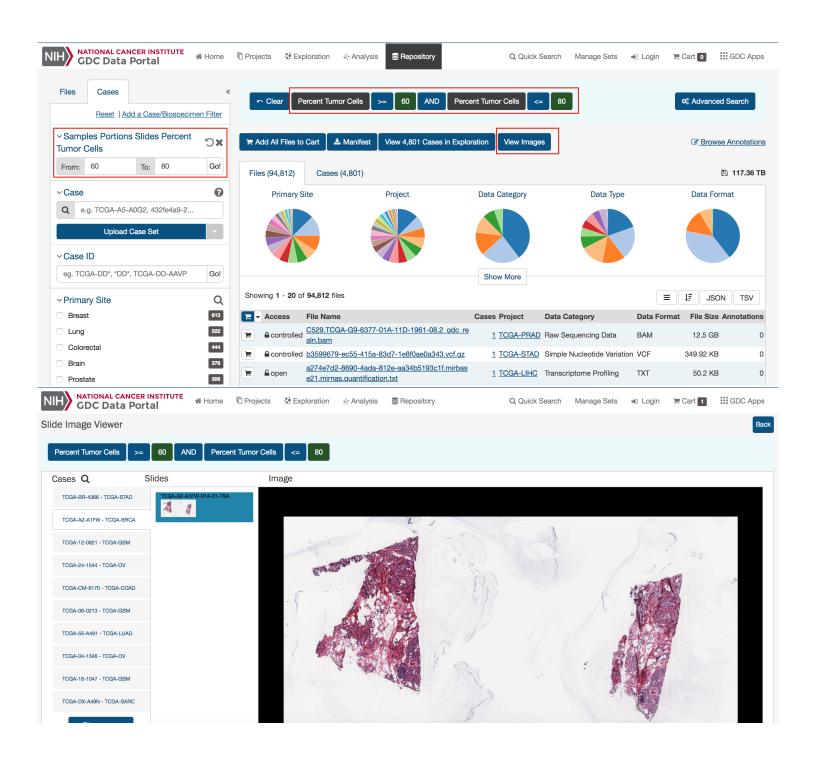
In the image viewer, you can: * Zoom in and zoom out by clicking on + and - icons * Reset to default display by clicking on the Home icon * Display the image in full screen mode by clicking on the Expand icon * View the slide detail by clicking on "Details" button * Selecting the area of interest with the thumbnail at the top-right corner



Example of navigation to the image viewer

- 1. Go to Repository Case facet and click on Add a Case/Biospecimen filter
- 2. In the search box of the filter pop-up, look to percent_tumor_cells and add this filter to the Repository
- 3. Enter a percentage between 60% and 80% then click on Go!
- 4. Click on "View Images" button

Result: The images displayed on the image viewer are filtered based on your query.



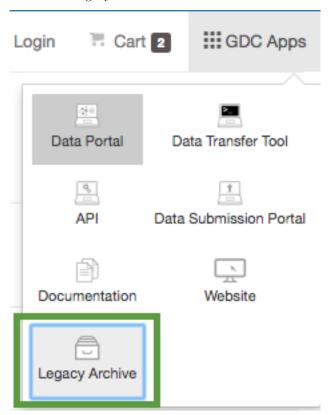
Chapter 12

Legacy Archive

Legacy Archive

The GDC Legacy Archive hosts unharmonized legacy data from repositories that predate the GDC (e.g. CGHub). Legacy data is not actively maintained, processed, or harmonized by the GDC. Legacy users are encouraged to migrate to harmonized datasets.

The GDC Legacy Archive can be accessed from the GDC Data Portal front page as well as from the "GDC Apps" menu.

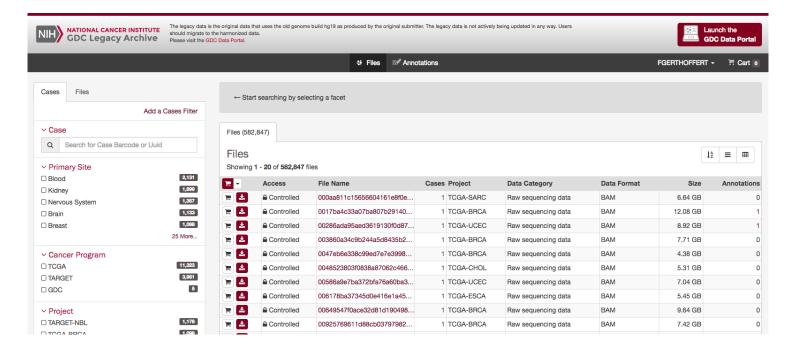


Overview

The GDC Legacy Archive contains a limited set of features of the GDC Data Portal:

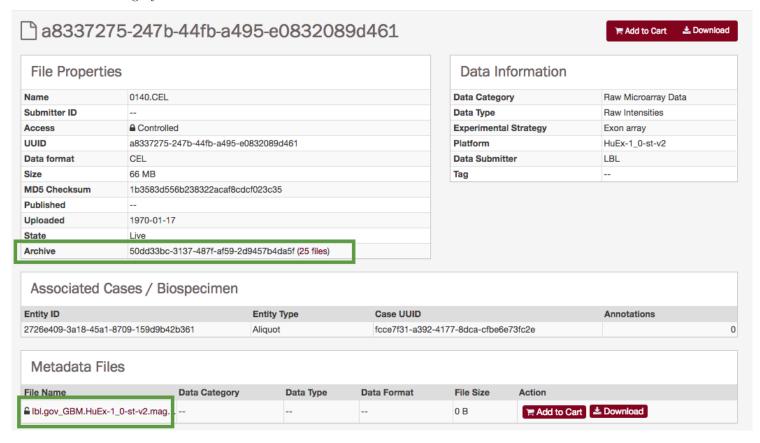
- Facet search: Ability to look for legacy files or legacy annotations based on case, file and annotation facets.
- File and Annotation tables: List of all the legacy files and list of all the legacy annotations.
- File and Annotation detail pages: Information page for each legacy file and annotation.

• Cart: The GDC Legacy Archive and the GDC Data Portal are separate systems with separate download carts.



File Page

The file page of the GDC Legacy Archive is similar to the File Summary Page of the GDC Data Portal. It does not include the Workflow, Reference Genome, and Read Groups sections as these are only applicable to harmonized data available in the GDC Data Portal. The Legacy Archive includes additional archive information as described below.



Archive

If a file was originally produced as part of an archive containing other files, the archive information (Archive ID and number of files in the archive) is displayed in the file properties and, if selected, the user will see a list of files containing all other files in that archive.

Metadata Files

If a file has any associated MAGE-TAB or SRA XML metadata files, these files will be listed at the bottom of the page. These files will can be downloaded directly from this page. Alternatively, metadata files can be downloaded from the file cart.

File Cart

The file cart in the GDC Legacy Archive is analogous to the file cart of the GDC Data Portal. It provides an additional button to download any SRA-XML and MAGE-TAB metadata files associated with the files in the cart.

Chapter 13

Release Notes

Data Portal Release Notes

5.1 5.0 4.1	Date
	August 14, 2020
4.1	July 2, 2020
1.1	March 10, 2020
3.1	December 10, 2019
3.0	November 6, 2019
2.0	July 31, 2019
1.0	June 5, 2019
0.0	April 17, 2019
9.0	February 20, 2019
8.0	December 18, 2018
7.0	November 7, 2018
6.0	September 27, 2018
5.0	August 23, 2018
4.0	June 13, 2018
3.0	May 21, 2018
2.0	February 15, 2018
1.0	December 21, 2017
0.0	November 16, 2017
0.0	October 24, 2017
.0	August 22, 2017
0.0	June 29, 2017
.2	May 9, 2017
.1	October 31, 2016
.0	September 7, 2016
1.0	August 9th, 2016
.0	June 1st, 2016

Version	Date
v1.0.1	May 18, 2016

Release 1.25.1

GDC Product: GDC Data Portal
Release Date: August 14, 2020

New Features and Changes

• API improvements were made to increase portal performance.

Bugs Fixed Since Last Release

• None

- Pre-release Data Portal login is not supported on Internet Explorer or the last version of Edge (42). Edge 41 does login successfully.
- Custom Facet Filters
 - Some definitions are missing from the property list when adding custom facet file or case filters.
- Visualizations
 - Negative numbers may be displayed for the Missing value category in the Treatment node within a Clinical Analysis.
 This occurs with projects that have multiple treatment nodes per case. All other values should be accurate.
 - SIFT and PolyPhen annotations are missing from the export JSON of the mutation table. They are present in the export TSV.
 - Data Portal graphs cannot be exported as PNG images in Internet Explorer. Graphs can be exported in PNG or SVG format from Chrome or Firefox browsers. Internet Explorer does not display chart legend and title when re-opening previously downloaded SVG files, the recommendation is to open downloaded SVG files with another program.
- Repository and Cart
 - The annotation count in File table of Repository and Cart does not link to the Annotations page anymore. The user can navigate to the annotations through the annotation count in Repository Case table.
- Legacy Archive
 - The footer says version 1.9, but it is actually 1.13
 - Filtering by vital_status does not function in the Legacy Archive due to updates in how this property has been indexed.
 A workaround is to perform the case level filtering in the GDC Data Portal and copy the filter string for use in the Legacy Archive or the legacy API.
 - Downloading a token in the GDC Legacy Archive does not refresh it. If a user downloads a token in the GDC Data Portal and then attempts to download a token in the GDC Legacy Archive, an old token may be provided. Reloading the Legacy Archive view will allow the user to download the updated token.
 - Exporting the Cart table in JSON will export the GDC Archive file table instead of exporting the files in the Cart only.
- Web Browsers

- Browsers limit the number of concurrent downloads, it is generally recommended to add files to the cart and download large number of files through the GDC Data Transfer Tool, more details can be found on GDC Website.
- The GDC Portals are not compatible with Internet Explorer running in compatibility mode. Workaround is to disable compatibility mode.

Release 1.25.0

• Release Date: July 2, 2020

New Features and Changes

- Suppressed Experimental Strategy filter on the Exploration page as this currently filters for files with a particular strategy, not for cases. This may cause confusion amongst users. The filter will be re-instated in a future release once the logic is available to filter more appropriately for cases tied to a specific strategy.
- Updated the filter control panel styling across the Portal to have clearer titles (e.g. "Search Cases" instead of "Cases" in the quick search box).
- Made minor updates to the styling of the filter query display at the top of the Exploration page (spacing, borders).
- Added an expand/collapse control to the quick search bar of Clinical tab on the Exploration page, to be consistent with other Exploration tabs.
- Added a clear title above the counts in each filter control panel across the Portal (e.g. " Cases", " Genes", etc.).
- Moved various action buttons above the results table on the Repository Page to more accessible locations.
- Improved load time of the initial custom filter list on the Repository Page, when clicking "Add a Filter Filter" or "Add a Case/Biospecimen Filter".

Bugs Fixed Since Last Release

- Fixed a bug in the Age at Diagnosis table on the Cohort Comparison page, where the of cases in the table was not consistent with the of cases shown when clicking the link to the Exploration page.
- Fixed minor positional accuracy issue of the lollipop data points on the Protein Viewer.
- Fixed bug on the Protein Viewer where, if clicking to switch between different lollipop data points, details of the previous lollipop was not closing.
- Fixed bug where the quick search bar on the Exploratin Page's Genes filter tab was not expanding/collapsing properly.
- Fixed bug in the pop-up warning message when adding or removing items from the Cart, where long filenames were spilling outside the border of the pop-up.
- Fixed typo in the "View Cases in Exploration" button on the Repository page.
- Fixed typo in the pop-up user consent message when downloading controlled files from the Cart.

- Pre-release Data Portal login is not supported on Internet Explorer or the last version of Edge (42). Edge 41 does login successfully.
- Custom Facet Filters
 - Some definitions are missing from the property list when adding custom facet file or case filters.
- Visualizations
 - Negative numbers may be displayed for the Missing value category in the Treatment node within a Clinical Analysis.
 This occurs with projects that have multiple treatment nodes per case. All other values should be accurate.
 - SIFT and PolyPhen annotations are missing from the export JSON of the mutation table. They are present in the export TSV.

– Data Portal graphs cannot be exported as PNG images in Internet Explorer. Graphs can be exported in PNG or SVG format from Chrome or Firefox browsers. Internet Explorer does not display chart legend and title when re-opening previously downloaded SVG files, the recommendation is to open downloaded SVG files with another program.

• Repository and Cart

- The annotation count in File table of Repository and Cart does not link to the Annotations page anymore. The user can navigate to the annotations through the annotation count in Repository - Case table.

• Legacy Archive

- The footer says version 1.9, but it is actually 1.13
- Filtering by vital_status does not function in the Legacy Archive due to updates in how this property has been indexed.
 A workaround is to perform the case level filtering in the GDC Data Portal and copy the filter string for use in the Legacy Archive or the legacy API.
- Downloading a token in the GDC Legacy Archive does not refresh it. If a user downloads a token in the GDC Data Portal and then attempts to download a token in the GDC Legacy Archive, an old token may be provided. Reloading the Legacy Archive view will allow the user to download the updated token.
- Exporting the Cart table in JSON will export the GDC Archive file table instead of exporting the files in the Cart only.

• Web Browsers

- Browsers limit the number of concurrent downloads, it is generally recommended to add files to the cart and download large number of files through the GDC Data Transfer Tool, more details can be found on GDC Website.
- The GDC Portals are not compatible with Internet Explorer running in compatibility mode. Workaround is to disable compatibility mode.

Release 1.24.1

GDC Product: GDC Data Portal
Release Date: March 10, 2020

New Features and Changes

- Removed unnecessary comma and y-axis value from title of the mutation details pop-up in the Protein Viewer.
- Added Tobacco Smoking Status field to the Exposures tab on the Case entity page.
- Added a link to the Cart where users can access instructions for downloading the GDC Genome Build reference files.
- Added logic to prevent duplicate fetching of data for Clinical Analysis survival plots and optimize rendering.
- Added a button to clear searches for certain Portal search controls that were previously missing this ability.
- Reduced whitespace between Oncogrid and its control panel to optimize spacing and layout.
- Made entire Clinical Analysis results page responsive (card columns now scale & stack in response to the size of the browser window).
- Replaced Clinical Analysis function for printing clinical cards to a single PDF file, with more flexible functionality to instead download all the cards in SVG and/or PNG format.
- Added message to notify users when they try to access the Portal using Microsoft Internet Explorer, indicating which browsers are officially supported.
- Added arrow icon to sortable columns across the Portal to indicate the current sort direction.

Bugs Fixed Since Last Release

- Fixed bug where clicking a primary site on the Human Body Image was not re-directing to the Exploration page.
- Fixed layout issue where long Annotation Notes were exceeding the border of the text box.
- Fixed layout issue where the Repository header and action buttons were scaling and wrapping incorrectly if the browser window is shrunk beyond a certain threshold.
- Fixed layout issue where the responsive Clinical Analysis Cards were clipping improperly as the browser window is shrunk beyond a certain threshold.

- Fixed bug where the Clinical Tab on the Exploration page was crashing when entering a custom range of Years for the Age at Diagnosis facet.
- Fixed various minor cosmetic and color issues in PNG, SVG downloads of the Clinical Analysis survival plots.
- Fixed bug where the x-axis in PNG, SVG downloads of histograms across the Portal was being bolded incorrectly.
- Fixed bug where the expand/collapse symbols in the UI were incorrectly being exported in the TSV download of the Projects table.
- Fixed bug where Oncogrid's modal for customizing colors could not be scrolled below the fold if it was shrunk beyond a
 certain threshold.
- Fixed incorrect DTT hyperlink in the GDC Apps menu.
- Fixed bug where the "dbSNP rs ID" facet could not be minimized in the Exploration page's Mutations facet tab.
- Fixed layout issue where the Portal's header incorrectly overlaps some content when a notification banner is displayed.
- Fixed some minor layout & styling issues in the Exploration page's facets panel.
- Fixed bug where the Case ID on the Exploration page's Cases facet tab was not searchable in certain scenarios.
- Fixed bug where the Expand/Collapse button state was not changing properly when being used in the Biospecimen section of the Case entity page.
- Fixed incorrect capitalization of "dbGaP" in the Summary section of the Project entity page.
- Fixed layout issue where the Advanced Search query box on the Repository page could expanded beyond the margins of the box's border.

Known Issues and Workarounds

- Pre-release Data Portal login is not supported on Internet Explorer or the last version of Edge (42). Edge 41 does login successfully.
- Custom Facet Filters
 - Some definitions are missing from the property list when adding custom facet file or case filters.

• Visualizations

- Negative numbers may be displayed for the Missing value category in the Treatment node within a Clinical Analysis.
 This occurs with projects that have multiple treatment nodes per case. All other values should be accurate.
- SIFT and PolyPhen annotations are missing from the export JSON of the mutation table. They are present in the export TSV.
- Data Portal graphs cannot be exported as PNG images in Internet Explorer. Graphs can be exported in PNG or SVG format from Chrome or Firefox browsers. Internet Explorer does not display chart legend and title when re-opening previously downloaded SVG files, the recommendation is to open downloaded SVG files with another program.

• Repository and Cart

- The annotation count in File table of Repository and Cart does not link to the Annotations page anymore. The user can navigate to the annotations through the annotation count in Repository - Case table.

• Legacy Archive

- The footer says version 1.9, but it is actually 1.13
- Filtering by vital_status does not function in the Legacy Archive due to updates in how this property has been indexed.
 A workaround is to perform the case level filtering in the GDC Data Portal and copy the filter string for use in the Legacy Archive or the legacy API.
- Downloading a token in the GDC Legacy Archive does not refresh it. If a user downloads a token in the GDC Data Portal and then attempts to download a token in the GDC Legacy Archive, an old token may be provided. Reloading the Legacy Archive view will allow the user to download the updated token.
- Exporting the Cart table in JSON will export the GDC Archive file table instead of exporting the files in the Cart only.

• Web Browsers

- Browsers limit the number of concurrent downloads, it is generally recommended to add files to the cart and download large number of files through the GDC Data Transfer Tool, more details can be found on GDC Website.
- The GDC Portals are not compatible with Internet Explorer running in compatibility mode. Workaround is to disable compatibility mode.

Release 1.23.1

GDC Product: GDC Data Portal
Release Date: December 10, 2019

New Features and Changes

• Updated display of x-axis units on the homepage Human Body chart to more easily display increased case counts for newly-added projects

Bugs Fixed Since Last Release

• None

Known Issues and Workarounds

- Pre-release Data Portal login is not supported on Internet Explorer or the last version of Edge (42). Edge 41 does login successfully.
- Custom Facet Filters
 - Some definitions are missing from the property list when adding custom facet file or case filters.
- Visualizations
 - Negative numbers may be displayed for the Missing value category in the Treatment node within a Clinical Analysis.
 This occurs with projects that have multiple treatment nodes per case. All other values should be accurate.
 - SIFT and PolyPhen annotations are missing from the export JSON of the mutation table. They are present in the export TSV.
 - Data Portal graphs cannot be exported as PNG images in Internet Explorer. Graphs can be exported in PNG or SVG format from Chrome or Firefox browsers. Internet Explorer does not display chart legend and title when re-opening previously downloaded SVG files, the recommendation is to open downloaded SVG files with another program.
- Repository and Cart
 - The annotation count in File table of Repository and Cart does not link to the Annotations page anymore. The user can navigate to the annotations through the annotation count in Repository Case table.
- Legacy Archive
 - The footer says version 1.9, but it is actually 1.13
 - Filtering by vital_status does not function in the Legacy Archive due to updates in how this property has been indexed.
 A workaround is to perform the case level filtering in the GDC Data Portal and copy the filter string for use in the Legacy Archive or the legacy API.
 - Downloading a token in the GDC Legacy Archive does not refresh it. If a user downloads a token in the GDC Data Portal and then attempts to download a token in the GDC Legacy Archive, an old token may be provided. Reloading the Legacy Archive view will allow the user to download the updated token.
 - Exporting the Cart table in JSON will export the GDC Archive file table instead of exporting the files in the Cart only.
- Web Browsers
 - Browsers limit the number of concurrent downloads, it is generally recommended to add files to the cart and download large number of files through the GDC Data Transfer Tool, more details can be found on GDC Website.
 - The GDC Portals are not compatible with Internet Explorer running in compatibility mode. Workaround is to disable compatibility mode.

Release 1.23.0

GDC Product: GDC Data Portal
Release Date: November 6, 2019

New Features and Changes

- Added Clinical Data Analysis feature that allows Users to:
 - Explore clinical data via the new Clinical Tab on the Exploration page.
 - Build custom Case sets based on that clinical data for later analysis.
 - Create an analysis to examine the clinical variables in a Case set, using various tools including histograms, survival plots, box plots, QQ plots, and custom binning.
 - Download the data (as TSV, JSON) and plots (as PNG, SVG) of each clinical variable in an anlysis.
 - Save an analysis to local storage to resume later (as long as storage is not cleared).
- Added links to CIViC annotations on the Gene and Mutation entity pages.
- Updated the default Top Mutated Genes histogram on the Exploration page to display only COSMIC Genes by default.
- Added Follow-Ups tab and nested Molecular Tests to Case entity page.
- Added text to BAM slicing modal to instruct Users how to access unmapped reads.

Bugs Fixed Since Last Release

- Fixed font in exported PNGs, SVGs to be consistent with the Portal UI.
- Made custom Case and File filters in the Repository page case insensitive.
- Fixed bug where pfam domains in Protein Viewer could not be clicked in Firefox.
- Fixed bug where TSV download button could not be clicked in MS Edge.
- Fixed controlled access alert pop-up in the Cart so that the modal disappears correctly once the User has successfully logged in and initiated the download.

- Pre-release Data Portal login is not supported on Internet Explorer or the last version of Edge (42). Edge 41 does login successfully.
- Custom Facet Filters
 - Some definitions are missing from the property list when adding custom facet file or case filters.
- Visualizations
 - Negative numbers may be displayed for the Missing value category in the Treatment node within a Clinical Analysis.
 This occurs with projects that have multiple treatment nodes per case. All other values should be accurate.
 - SIFT and PolyPhen annotations are missing from the export JSON of the mutation table. They are present in the export TSV.
 - Data Portal graphs cannot be exported as PNG images in Internet Explorer. Graphs can be exported in PNG or SVG format from Chrome or Firefox browsers. Internet Explorer does not display chart legend and title when re-opening previously downloaded SVG files, the recommendation is to open downloaded SVG files with another program.
- Repository and Cart
 - The annotation count in File table of Repository and Cart does not link to the Annotations page anymore. The user can navigate to the annotations through the annotation count in Repository Case table.
- Legacy Archive
 - The footer says version 1.9, but it is actually 1.13
 - Filtering by vital_status does not function in the Legacy Archive due to updates in how this property has been indexed.
 A workaround is to perform the case level filtering in the GDC Data Portal and copy the filter string for use in the Legacy Archive or the legacy API.
 - Downloading a token in the GDC Legacy Archive does not refresh it. If a user downloads a token in the GDC Data Portal and then attempts to download a token in the GDC Legacy Archive, an old token may be provided. Reloading the Legacy Archive view will allow the user to download the updated token.
 - Exporting the Cart table in JSON will export the GDC Archive file table instead of exporting the files in the Cart only.
- Web Browsers

- Browsers limit the number of concurrent downloads, it is generally recommended to add files to the cart and download large number of files through the GDC Data Transfer Tool, more details can be found on GDC Website.
- The GDC Portals are not compatible with Internet Explorer running in compatibility mode. Workaround is to disable compatibility mode.

Release 1.22.0

GDC Product: GDC Data Portal
Release Date: July 31, 2019

New Features and Changes

- Replaced existing Clinical, Biospecimen columns on the Projects page with 4 columns: Clinical, Clinical Supplement, Biospecimen, Biospecimen Supplement. The Clinical and Biospecimen columns now link directly to the project page, and their counts indicate the total cases in the project. The Clinical Supplement and Biospecimen Supplement columns work the same as the old Clinical and Biospecimen columns They link to the Repository page with Files filtered based on the Project and Data Category (Clinical or Biospecimen).
- Added a new icon to the GDC Apps menu, which links to the GDC Publications website page.
- Added the Synchronous Malignancy field to the Diagnoses / Treatments tab on the Case entity page.
- Added the Pack Years Smoked field to the Exposures tab on the Case entity page.
- Increased length of x-axis labels on histograms to 10 characters so that projects with names that are typically standard 10 chars will display fully (e.g. most TCGA projects like TCGA-BRCA).

Bugs Fixed Since Last Release

• Fixed bug where the PNG, SVG files for the Overall Survival Plot could not be downloaded.

- Pre-release Data Portal login is not supported on Internet Explorer or the last version of Edge (42). Edge 41 does login successfully.
- Custom Facet Filters
 - Some definitions are missing from the property list when adding custom facet file or case filters.
- Visualizations
 - SIFT and PolyPhen annotations are missing from the export JSON of the mutation table. They are present in the export TSV.
 - Data Portal graphs cannot be exported as PNG images in Internet Explorer. Graphs can be exported in PNG or SVG format from Chrome or Firefox browsers. Internet Explorer does not display chart legend and title when re-opening previously downloaded SVG files, the recommendation is to open downloaded SVG files with another program.
- Repository and Cart
 - The annotation count in File table of Repository and Cart does not link to the Annotations page anymore. The user can navigate to the annotations through the annotation count in Repository - Case table.
- Legacy Archive
 - Filtering by vital_status does not function in the Legacy Archive due to updates in how this property has been indexed.
 A workaround is to perform the case level filtering in the GDC Data Portal and copy the filter string for use in the Legacy Archive or the legacy API.
 - Downloading a token in the GDC Legacy Archive does not refresh it. If a user downloads a token in the GDC Data Portal and then attempts to download a token in the GDC Legacy Archive, an old token may be provided. Reloading the Legacy Archive view will allow the user to download the updated token.

- Exporting the Cart table in JSON will export the GDC Archive file table instead of exporting the files in the Cart only.
- Web Browsers
 - Browsers limit the number of concurrent downloads, it is generally recommended to add files to the cart and download large number of files through the GDC Data Transfer Tool, more details can be found on GDC Website.
 - The GDC Portals are not compatible with Internet Explorer running in compatibility mode. Workaround is to disable compatibility mode.

Release 1.21.0

• GDC Product: GDC Data Portal

• Release Date: June 5, 2019

New Features and Changes

- Changed all Survival Plots to display the Duration (x-axis) in years instead of days.
- Updated data references to clinical properties throughout the Portal to match the underlying changes in the GDC data dictionary.

Bugs Fixed Since Last Release

- Fixed bug where X-axis labels in histograms were cut off when displayed.
- Renamed the 'Experimental Strategies' facet on the Projects page to singular form.
- Fixed bug where columns with a % value of infinity (due to division by zero) show as 'NaN%'. Replaced instead with a label of '-'.
- Fixed bug where the download button in the cart access banner was still disabled after a user logged in from the banner. Instead, the experience is now improved so that after login, the banner is closed and the user must explicitly click 'Download' again.
- Fixed bug where if a new user logs into the Portal and views their profile, the app crashes if the user has no projects assigned yet.
- Fixed bug where Survival Rate numbers in the Survival Plot plot y-axis did not scale properly and overlapped into the axis lines.

- Pre-release Data Portal login is not supported on Internet Explorer or the last version of Edge (42). Edge 41 does login successfully.
- Custom Facet Filters
 - Some definitions are missing from the property list when adding custom facet file or case filters.
- Visualizations
 - SIFT and PolyPhen annotations are missing from the export JSON of the mutation table. They are present in the export TSV.
 - Data Portal graphs cannot be exported as PNG images in Internet Explorer. Graphs can be exported in PNG or SVG format from Chrome or Firefox browsers. Internet Explorer does not display chart legend and title when re-opening previously downloaded SVG files, the recommendation is to open downloaded SVG files with another program.
- Repository and Cart
 - The annotation count in File table of Repository and Cart does not link to the Annotations page anymore. The user can navigate to the annotations through the annotation count in Repository Case table.
- Legacy Archive

- Downloading a token in the GDC Legacy Archive does not refresh it. If a user downloads a token in the GDC Data Portal and then attempts to download a token in the GDC Legacy Archive, an old token may be provided. Reloading the Legacy Archive view will allow the user to download the updated token.
- Exporting the Cart table in JSON will export the GDC Archive file table instead of exporting the files in the Cart only.

• Web Browsers

- Browsers limit the number of concurrent downloads, it is generally recommended to add files to the cart and download large number of files through the GDC Data Transfer Tool, more details can be found on GDC Website.
- The GDC Portals are not compatible with Internet Explorer running in compatibility mode. Workaround is to disable compatibility mode.

Release 1.20.0

GDC Product: GDC Data Portal
Release Date: April 17, 2019

New Features and Changes

• Upgraded the Portal to use the latest React Javascript library (version 16.8)

Bugs Fixed Since Last Release

• None

- Pre-release Data Portal login is not supported on Internet Explorer or the last version of Edge (42). Edge 41 does login successfully.
- Custom Facet Filters
 - Some definitions are missing from the property list when adding custom facet file or case filters.
- Visualizations
 - SIFT and PolyPhen annotations are missing from the export JSON of the mutation table. They are present in the export TSV.
 - Data Portal graphs cannot be exported as PNG images in Internet Explorer. Graphs can be exported in PNG or SVG format from Chrome or Firefox browsers. Internet Explorer does not display chart legend and title when re-opening previously downloaded SVG files, the recommendation is to open downloaded SVG files with another program.
- Repository and Cart
 - The annotation count in File table of Repository and Cart does not link to the Annotations page anymore. The user can navigate to the annotations through the annotation count in Repository Case table.
- Legacy Archive
 - Downloading a token in the GDC Legacy Archive does not refresh it. If a user downloads a token in the GDC Data Portal and then attempts to download a token in the GDC Legacy Archive, an old token may be provided. Reloading the Legacy Archive view will allow the user to download the updated token.
 - Exporting the Cart table in JSON will export the GDC Archive file table instead of exporting the files in the Cart only.
- Web Browsers
 - Browsers limit the number of concurrent downloads, it is generally recommended to add files to the cart and download large number of files through the GDC Data Transfer Tool, more details can be found on GDC Website.
 - The GDC Portals are not compatible with Internet Explorer running in compatibility mode. Workaround is to disable compatibility mode.

Release 1.19.0

GDC Product: GDC Data Portal
Release Date: February 20, 2019

New Features and Changes

- Added support for viewing of controlled-access mutations in the Data Portal
- Added a new data access notification to remind logged-in users with access to controlled data that they need to follow their data use agreement. The message is fixed at the top of the Portal.
- Added the ability to search for previous versions of files. If the user enters the UUID of a previous version that cannot be found, the Portal returns the UUID of the latest version available.
- Renamed the Data Category for "Raw Sequencing Data" to "Sequencing Reads" throughout the portal where this appears, to be consistent with the Data Dictionary.
- Added a link in the Portal footer to the GDC support page.

Bugs Fixed Since Last Release

- Fixed bug where Survival Plot button never stops loading if plotting mutated vs. non-mutated cases for a single Gene.
- Fixed inconsistent button styling when downloading controlled Downstream Analyses Files from File Entity page.
- Removed unnecessary Survival column from Arrange Columns button on Case Entity, Gene Entity pages.
- Removed unnecessary whitespace from pie charts on Repository page.
- Added missing File Size unit to Clinical Supplement File, Biospecimen Supplement File tables on Case Entity page.
- Fixed bug where clicking on Case Counts in Projects Graph tab was going to the Repository Files tab instead of the Cases tab.
- Fixed bug where the counts shown beside customer filters on the Repository Cases tab were not updating when filtering on other facets.
- Fixed bug where clicking the of Affected Cases denominator on the Gene page's Most Frequent Somatic Mutations table displayed an incorrect number of Cases.

- Pre-release Data Portal login is not supported on Internet Explorer or the last version of Edge (42). Edge 41 does login successfully.
- Custom Facet Filters
 - Some definitions are missing from the property list when adding custom facet file or case filters.
- Visualizations
 - SIFT and PolyPhen annotations are missing from the export JSON of the mutation table. They are present in the export TSV.
 - Data Portal graphs cannot be exported as PNG images in Internet Explorer. Graphs can be exported in PNG or SVG format from Chrome or Firefox browsers. Internet Explorer does not display chart legend and title when re-opening previously downloaded SVG files, the recommendation is to open downloaded SVG files with another program.
- Repository and Cart
 - The annotation count in File table of Repository and Cart does not link to the Annotations page anymore. The user can navigate to the annotations through the annotation count in Repository Case table.
- Legacy Archive
 - Downloading a token in the GDC Legacy Archive does not refresh it. If a user downloads a token in the GDC Data Portal and then attempts to download a token in the GDC Legacy Archive, an old token may be provided. Reloading the Legacy Archive view will allow the user to download the updated token.
 - Exporting the Cart table in JSON will export the GDC Archive file table instead of exporting the files in the Cart only.

• Web Browsers

- Browsers limit the number of concurrent downloads, it is generally recommended to add files to the cart and download large number of files through the GDC Data Transfer Tool, more details can be found on GDC Website.
- The GDC Portals are not compatible with Internet Explorer running in compatibility mode. Workaround is to disable compatibility mode.

Release 1.18.0

GDC Product: GDC Data Portal
Release Date: December 18, 2018

New Features and Changes

- A new data access message has been added when downloading controlled data. Users must agree to abide by data access control policies when downloading controlled data.
- In the Mutation free-text search in Exploration, mutation display now includes the UUID, genomic location, and matched search term for easier mutation searching.
- The ability to sort on ranked columns has been made available.

Bugs Fixed Since Last Release

- In some cases, text was being cut off on the Project page visualization tab. Text is no longer cut off.
- HGNC link on Gene page broke as the source format url changed; The format was updated and the link is now functional
- In the biospecimen details on the Case page, the cart icon would disappear once clicked. It now is always visible.

- Pre-release Data Portal login is not supported on Internet Explorer or the last version of Edge (42). Edge 41 does login successfully.
- Custom Facet Filters
 - Some definitions are missing from the property list when adding custom facet file or case filters.
- Visualizations
 - SIFT and PolyPhen annotations are missing from the export JSON of the mutation table. They are present in the export TSV.
 - Data Portal graphs cannot be exported as PNG images in Internet Explorer. Graphs can be exported in PNG or SVG format from Chrome or Firefox browsers. Internet Explorer does not display chart legend and title when re-opening previously downloaded SVG files, the recommendation is to open downloaded SVG files with another program.
- Repository and Cart
 - The annotation count in File table of Repository and Cart does not link to the Annotations page anymore. The user can navigate to the annotations through the annotation count in Repository Case table.
- Legacy Archive
 - Downloading a token in the GDC Legacy Archive does not refresh it. If a user downloads a token in the GDC Data Portal and then attempts to download a token in the GDC Legacy Archive, an old token may be provided. Reloading the Legacy Archive view will allow the user to download the updated token.
 - Exporting the Cart table in JSON will export the GDC Archive file table instead of exporting the files in the Cart only.
- Web Browsers
 - Browsers limit the number of concurrent downloads, it is generally recommended to add files to the cart and download large number of files through the GDC Data Transfer Tool, more details can be found on GDC Website.
 - The GDC Portals are not compatible with Internet Explorer running in compatibility mode. Workaround is to disable compatibility mode.

Release 1.17.0

GDC Product: GDC Data Portal
Release Date: November 7, 2018

New Features and Changes

- Copy Number Variation (CNV) data derived from GISTIC results are now available in the portal:
 - View number of CNV events on a gene in a cohort in the Explore Gene table tab
 - Explore CNVs associated with a gene on the Gene Entity Page
 - Explore CNVs concurrently with mutations on the Oncogrid with new visualization

Bugs Fixed Since Last Release

• None

Known Issues and Workarounds

- Custom Facet Filters
 - Some definitions are missing from the property list when adding custom facet file or case filters.
- Visualizations
 - SIFT and PolyPhen annotations are missing from the export JSON of the mutation table. They are present in the export TSV.
 - Data Portal graphs cannot be exported as PNG images in Internet Explorer. Graphs can be exported in PNG or SVG format from Chrome or Firefox browsers. Internet Explorer does not display chart legend and title when re-opening previously downloaded SVG files, the recommendation is to open downloaded SVG files with another program.
- Repository and Cart
 - The annotation count in File table of Repository and Cart does not link to the Annotations page anymore. The user can navigate to the annotations through the annotation count in Repository Case table.
- Legacy Archive
 - Downloading a token in the GDC Legacy Archive does not refresh it. If a user downloads a token in the GDC Data Portal and then attempts to download a token in the GDC Legacy Archive, an old token may be provided. Reloading the Legacy Archive view will allow the user to download the updated token.
 - Exporting the Cart table in JSON will export the GDC Archive file table instead of exporting the files in the Cart only.
- Web Browsers
 - Browsers limit the number of concurrent downloads, it is generally recommended to add files to the cart and download large number of files through the GDC Data Transfer Tool, more details can be found on GDC Website.
 - The GDC Portals are not compatible with Internet Explorer running in compatibility mode. Workaround is to disable compatibility mode.

Release 1.16.0

GDC Product: GDC Data Portal
Release Date: September 27, 2018

New Features and Changes

• Updated Human Body Image to aggregate all current primary sites to available Major Primary Sites

- Fixed link on cart download error popup
- Updated Cancer Distribution table to have dropdown menus for primary site and disease type
- Updated Y-axis label on Top Mutated Cancer Genes in Selected Projects Graph
- Updated Set Operation Image to remove stray text

Known Issues and Workarounds

- Advanced Search
 - For advanced search and custom file facet filtering there are some properties that will appear as options that are no longer supported (e.g. file_state).
- Custom facet filters
 - Some definitions are missing from the property list when adding custom facet file or case filters.
- Visualizations
 - SIFT and PolyPhen annotations are missing from the Export JSON of the mutation table. They are present in the export TSV.
 - Data Portal graphs cannot be exported as PNG images in Internet Explorer. Graphs can be exported in PNG or SVG format from Chrome or Firefox browsers. Internet Explorer does not display chart legend and title when re-opening previously downloaded SVG files, the recommendation is to open downloaded SVG files with another program.
- Repository and Cart
 - The annotation count in File table of Repository and Cart does not link to the Annotations page anymore. The user can navigate to the annotations through the annotation count in Repository Case table.
- Legacy Archive
 - Downloading a token in the GDC Legacy Archive does not refresh it. If a user downloads a token in the GDC Data Portal and then attempts to download a token in the GDC Legacy Archive, an old token may be provided. Reloading the Legacy Archive view will allow the user to download the updated token.
 - Exporting the Cart table in JSON will export the GDC Archive file table instead of exporting the files in the Cart only.
- Web Browsers
 - Browsers limit the number of concurrent downloads, it is generally recommended to add files to the cart and download large number of files through the GDC Data Transfer Tool, more details can be found on GDC Website.
 - The GDC Portals are not compatible with Internet Explorer running in compatibility mode. Workaround is to disable compatibility mode.

Release 1.15.0

GDC Product: GDC Data Portal
Release Date: August 23, 2018

New Features and Changes

- File Versions are now visible in the "File Versions" section on the File Entity Page.
- "View Files in Repository" and "View Cases in Repository" button methods were updated to work faster.

- Fixed warning messages that prompted users to login even when already logged in. Error warnings now correctly prompt users to reference dbGAP for data access if already signed in.
- Fixed error where you could click Go on Case ID wildcard facet before inputting any data.
- Fixed cart header to be a consistent color for the whole table.
- Fixed error where you could save a set with no name or items, which resulted in an infinite spinner.
- Fixed table width issue when FM-AD was selected as a filter.
- Updated broken help link on Advanced Query.

Known Issues and Workarounds

- Advanced Search
 - For advanced search and custom file facet filtering there are some properties that will appear as options that are no longer supported (e.g. file_state).
- Custom facet filters
 - Some definitions are missing from the property list when adding custom facet file or case filters.
- Visualizations
 - SIFT and PolyPhen annotations are missing from the Export JSON of the mutation table. They are present in the export TSV.
 - Data Portal graphs cannot be exported as PNG images in Internet Explorer. Graphs can be exported in PNG or SVG format from Chrome or Firefox browsers. Internet Explorer does not display chart legend and title when re-opening previously downloaded SVG files, the recommendation is to open downloaded SVG files with another program.
- Repository and Cart
 - The annotation count in File table of Repository and Cart does not link to the Annotations page anymore. The user can navigate to the annotations through the annotation count in Repository Case table.
- Legacy Archive
 - Downloading a token in the GDC Legacy Archive does not refresh it. If a user downloads a token in the GDC Data Portal and then attempts to download a token in the GDC Legacy Archive, an old token may be provided. Reloading the Legacy Archive view will allow the user to download the updated token.
 - Exporting the Cart table in JSON will export the GDC Archive file table instead of exporting the files in the Cart only.
- Web Browsers
 - Browsers limit the number of concurrent downloads, it is generally recommended to add files to the cart and download large number of files through the GDC Data Transfer Tool, more details can be found on GDC Website.
 - The GDC Portals are not compatible with Internet Explorer running in compatibility mode. Workaround is to disable compatibility mode.

Release 1.14.0

GDC Product: GDC Data Portal
Release Date: June 13, 2018

New Features and Changes

 Added new Experimental Strategies Diagnostic Slide Image, Bisulfite-Seq, ChIP-Seq, and ATAC-Seq to Case and Project entity pages.

• Fixed download of clinical and biospecimen data from the Repository when Case table rows are selected.

Known Issues and Workarounds

- Custom facet filters
 - Some definitions are missing from the property list when adding custom facet file or case filters.
- Visualizations
 - SIFT and PolyPhen annotations are missing from the Export JSON of the mutation table. They are present in the export TSV.
 - Data Portal graphs cannot be exported as PNG images in Internet Explorer. Graphs can be exported in PNG or SVG format from Chrome or Firefox browsers. Internet Explorer does not display chart legend and title when re-opening previously downloaded SVG files, the recommendation is to open downloaded SVG files with another program.
- Repository and Cart
 - When user is logged in and try to download a controlled file he does not have access to, he's prompted to log in. He should be promted to request access.
 - The annotation count in File table of Repository and Cart does not link to the Annotations page anymore. The user can navigate to the annotations through the annotation count in Repository Case table.
- Legacy Archive
 - Downloading a token in the GDC Legacy Archive does not refresh it. If a user downloads a token in the GDC Data Portal and then attempts to download a token in the GDC Legacy Archive, an old token may be provided. Reloading the Legacy Archive view will allow the user to download the updated token.
 - Exporting the Cart table in JSON will export the GDC Archive file table instead of exporting the files in the Cart only.
- Web Browsers
 - Browsers limit the number of concurrent downloads, it is generally recommended to add files to the cart and download large number of files through the GDC Data Transfer Tool, more details can be found on GDC Website.
 - The GDC Portals are not compatible with Internet Explorer running in compatibility mode. Workaround is to disable compatibility mode.

Release details are maintained in the GDC Data Portal Change Log.

Release 1.13.0

GDC Product: GDC Data Portal
Release Date: May 21, 2018

New Features and Changes

• Added new image viewer functionality for viewing tissue slide images

Bugs Fixed Since Last Release

- Updated gene reference labels on gene entity page to adhere to preferred usage
- Fixed issue with user profile displaying all projects twice

- Custom facet filters
 - Some definitions are missing from the property list when adding custom facet file or case filters.
- Visualizations
 - SIFT and PolyPhen annotations are missing from the Export JSON of the mutation table. They are present in the export TSV.
 - Data Portal graphs cannot be exported as PNG images in Internet Explorer. Graphs can be exported in PNG or SVG format from Chrome or Firefox browsers. Internet Explorer does not display chart legend and title when re-opening previously downloaded SVG files, the recommendation is to open downloaded SVG files with another program.
- Repository and Cart
 - When user is logged in and try to download a controlled file he does not have access to, he's prompted to log in. He should be promted to request access.
 - The annotation count in File table of Repository and Cart does not link to the Annotations page anymore. The user can navigate to the annotations through the annotation count in Repository Case table.
- Legacy Archive
 - Downloading a token in the GDC Legacy Archive does not refresh it. If a user downloads a token in the GDC Data
 Portal and then attempts to download a token in the GDC Legacy Archive, an old token may be provided. Reloading
 the Legacy Archive view will allow the user to download the updated token.
 - Exporting the Cart table in JSON will export the GDC Archive file table instead of exporting the files in the Cart only.
- Web Browsers
 - Browsers limit the number of concurrent downloads, it is generally recommended to add files to the cart and download large number of files through the GDC Data Transfer Tool, more details can be found on GDC Website.
 - The GDC Portals are not compatible with Internet Explorer running in compatibility mode. Workaround is to disable compatibility mode.

Release details are maintained in the GDC Data Portal Change Log.

Release 1.12.0

GDC Product: GDC Data Portal
Release Date: February 15, 2018

New Features and Changes

- Provided the ability to export clinical and biospecimen data in a TSV format from the Case, Project, Exploration, Repository and Cart pages.
- Removed from the Project entity page the sections about mutated genes, somatic mutations and affected Cases and replaced with a button "Explore data" that will open the Exploration page filtered on the project. Indeed the Exploration page provides the same information. Added a breakdown of cases per primary site for a Project entity page with multiple primary sites (e.g. FM-AD).
- Added display of coding DNA change and impacts for all the transcripts (instead of canonical transcript only) in the Mutation entity page Consequences section. In the mutation table (e.g. in Repository), the impacts and consequences are displayed for the canonical transcript only.

Bugs Fixed Since Last Release

 Replaced the suggested set name when saving a set with selected items, e.g. for case set the suggested name is now "Custom Case selection".

- Fixed the protein viewer to indicate when there are overlapping mutations. Mousing over the dot showing multiple mutations will open a right panel with the list of all the corresponding mutations.
- Fixed Mutation entity page Consequences table: the "Coding DNA Change" column is now populated for all the transcripts.
- Fixed download clinical and download biospecimen actions from TCGA-BRCA project.
- Fixed facet behavior that did not reset back to showing all options after pressing reset-arrow.
- Fixed error when user was trying to save a set with no value in the textbox "Save top:".
- Removed somatic mutation section from Case entity page for cases with no open-access mutation data (e.g. FM-AD or TARGET cases).
- Fixed error where a blank page appears after unselecting Cancer Gene Census mutation facet.
- Fixed duplicated date in sample sheet name (e.g. gdc_sample_sheet_YYYY-MM-DD_HH-MM.tsv.YYYY-MM-DD_HH-MM.tsv).
- Fixed error when annotations were not downloaded along with the file (in File entity page and Cart).

- Custom facet filters
 - Some definitions are missing from the property list when adding custom facet file or case filters.
- Visualizations
 - SIFT and PolyPhen annotations are missing from the Export JSON of the mutation table. They are present in the export TSV.
 - Data Portal graphs cannot be exported as PNG images in Internet Explorer. Graphs can be exported in PNG or SVG format from Chrome or Firefox browsers. Internet Explorer does not display chart legend and title when re-opening previously downloaded SVG files, the recommendation is to open downloaded SVG files with another program.
- Repository and Cart
 - When user is logged in and try to download a controlled file he does not have access to, he's prompted to log in. He should be promted to request access.
 - The annotation count in File table of Repository and Cart does not link to the Annotations page anymore. The user can navigate to the annotations through the annotation count in Repository Case table.
- · Legacy Archive
 - Downloading a token in the GDC Legacy Archive does not refresh it. If a user downloads a token in the GDC Data
 Portal and then attempts to download a token in the GDC Legacy Archive, an old token may be provided. Reloading
 the Legacy Archive view will allow the user to download the updated token.
 - Exporting the Cart table in JSON will export the GDC Archive file table instead of exporting the files in the Cart only.
- Web Browsers
 - Browsers limit the number of concurrent downloads, it is generally recommended to add files to the cart and download large number of files through the GDC Data Transfer Tool, more details can be found on GDC Website.
 - The GDC Portals are not compatible with Internet Explorer running in compatibility mode. Workaround is to disable compatibility mode.

Release details are maintained in the GDC Data Portal Change Log.

Release 1.11.0

GDC Product: GDC Data Portal
Release Date: December 21, 2017

New Features and Changes

- Updated UI to support SIFT and Polyphen annotations
- A Sample Sheet can now be created which allows easy association between file names and the case and sample submitter_id
- Updated Advanced Search page to include options to Add All Files to Cart, Download Manifest, and View X Cases in Exploration
- Provide clear message rather than blank screen if survival plots cannot be calculated for particular cohort comparison
- Display sample_type on associated entities section on file page
- Allows for special characters in case, gene, and mutation set upload (-, :, >, .)

Bugs Fixed Since Last Release

- Fixed error when trying to download large number of files from the Legacy Archive cart
- Fixed number of annotations displayed in Legacy Archive for particular entities
- Replaced missing bars to indicate proportion of applicable files and cases on project entity page in Cases and File Counts by Data Category table
- Fixed project page display when projects are selected that contain no mutation data in the facet panel
- Fixed error where exporting case sets as TSV included fewer cases than the total
- Fixed error in exploration section when adding custom facets. Previously selecting 'Only show fields with values' did not result in the expected behavior
- Fixed error where number of associated entities for a file was showing an incorrect number

Known Issues and Workarounds

- Sample sheet will download with a file name including the date duplicated (e.g. gdc_sample_sheet_YYYY-MM-DD_HH-MM.tsv.YYYY-MM-DD HH-MM.tsv)
- Custom facet filters
 - Definitions are missing from the property list when adding custom facet file or case filters
- Visualizations
 - Data Portal graphs cannot be exported as PNG images in Internet Explorer. Graphs can be exported in PNG or SVG format from Chrome or Firefox browsers. Internet Explorer does not display chart legend and title when re-opening previously downloaded SVG files, the recommendation is to open downloaded SVG files with another program.
 - In the protein viewer there may be overlapping mutations. In this case mousing over a point will just show a single mutation and the other mutations at this location will not be apparent.
- Entity page
 - On the mutation entity page, in the Consequences Table, the "Coding DNA Change" column is not populated for rows
 that do not correspond to the canonical mutation.
- Repository and Cart
 - The annotation count in File table of Repository and Cart does not link to the Annotations page anymore. The user can navigate to the annotations through the annotation count in Repository Case table.
- · Legacy Archive
 - Downloading a token in the GDC Legacy Archive does not refresh it. If a user downloads a token in the GDC Data Portal and then attempts to download a token in the GDC Legacy Archive, an old token may be provided. Reloading the Legacy Archive view will allow the user to download the updated token.
 - Exporting the Cart table in JSON will export the GDC Archive file table instead of exporting the files in the Cart only.
- Web Browsers
 - Browsers limit the number of concurrent downloads, it is generally recommended to add files to the cart and download large number of files through the GDC Data Transfer Tool, more details can be found on GDC Website.
 - The GDC Portals are not compatible with Internet Explorer running in compatibility mode. Workaround is to disable compatibility mode.

Release 1.10.0

GDC Product: GDC Data Portal
Release Date: November 16, 2017

New Features and Changes

- Support for uploading Case and Mutation sets in Exploration page
- Support for saving, editing, removing Case, Gene and Mutation sets in the Exploration page
- Added a Managed Sets menu where the user can see their saved sets
- Added an Analysis menu with two analyses: Set Operation and Cohort Comparison
- Added a User Profile page that shows all the projects and permissions assigned to the user: available in the username dropdown after the user logs in

Bugs Fixed Since Last Release

- Project page
 - On the project page, the Summary Case Count link should open the case tab on the Repository page instead it opens the file page

Known Issues and Workarounds

- Custom facet filters
 - Definitions are missing from the property list when adding custom facet file or case filters
 - Selecting 'Only show fields with values' will show some fields without values in the Repository section. This works correctly under the Exploration section.
- Visualizations
 - Data Portal graphs cannot be exported as PNG images in Internet Explorer. Graphs can be exported in PNG or SVG format from Chrome or Firefox browsers. Internet Explorer does not display chart legend and title when re-opening previously downloaded SVG files, the recommendation is to open downloaded SVG files with another program.
 - In the protein viewer there may be overlapping mutations. In this case mousing over a point will just show a single mutation and the other mutations at this location will not be apparent.
- Entity page
 - On the mutation entity page, in the Consequences Table, the "Coding DNA Change" column is not populated for rows that do not correspond to the canonical mutation.
- Repository and Cart
 - The annotation count in File table of Repository and Cart does not link to the Annotations page anymore. The user can navigate to the annotations through the annotation count in Repository Case table.
- Legacy Archive
 - Downloading a token in the GDC Legacy Archive does not refresh it. If a user downloads a token in the GDC Data Portal and then attempts to download a token in the GDC Legacy Archive, an old token may be provided. Reloading the Legacy Archive view will allow the user to download the updated token.
 - Exporting the Cart table in JSON will export the GDC Archive file table instead of exporting the files in the Cart only.
- Web Browsers
 - Browsers limit the number of concurrent downloads, it is generally recommended to add files to the cart and download large number of files through the GDC Data Transfer Tool, more details can be found on GDC Website.
 - The GDC Portals are not compatible with Internet Explorer running in compatibility mode. Workaround is to disable compatibility mode.

Release 1.9.0

GDC Product: GDC Data Portal
Release Date: October 24, 2017

New Features and Changes

- Support for projects with multiple primary sites per project
- Support for slides that are linked to sample rather than portion

Bugs Fixed Since Last Release

None

Known Issues and Workarounds

- Visualizations
 - Data Portal graphs cannot be exported as PNG images in Internet Explorer. Graphs can be exported in PNG or SVG format from Chrome or Firefox browsers. Internet Explorer does not display chart legend and title when re-opening previously downloaded SVG files, the recommendation is to open downloaded SVG files with another program.
 - In the protein viewer there may be overlapping mutations. In this case mousing over a point will just show a single mutation and the other mutations at this location will not be apparent.
- Project page
 - On the project page, the Summary Case Count link should open the case tab on the Repository page instead it opens
 the file page
- Entity page
 - On the mutation entity page, in the Consequences Table, the "Coding DNA Change" column is not populated for rows that do not correspond to the canonical mutation.
- Repository and Cart
 - The annotation count in File table of Repository and Cart does not link to the Annotations page anymore. The user can navigate to the annotations through the annotation count in Repository Case table.
- Legacy Archive
 - Downloading a token in the GDC Legacy Archive does not refresh it. If a user downloads a token in the GDC Data Portal and then attempts to download a token in the GDC Legacy Archive, an old token may be provided. Reloading the Legacy Archive view will allow the user to download the updated token.
 - Exporting the Cart table in JSON will export the GDC Archive file table instead of exporting the files in the Cart only.
- Web Browsers
 - Browsers limit the number of concurrent downloads, it is generally recommended to add files to the cart and download large number of files through the GDC Data Transfer Tool, more details can be found on GDC Website.
 - The GDC Portals are not compatible with Internet Explorer running in compatibility mode. Workaround is to disable compatibility mode.

Release details are maintained in the GDC Data Portal Change Log.

Release 1.8.0

GDC Product: GDC Data Portal
Release Date: August 22, 2017

New Features and Changes

Major features/changes:

- A feature that links the exploration and repository pages was added. For example:
 - In the exploration page, cases with a specific mutation could be selected. This set could then be linked to the repository page to download the data files associated with these cases.
 - In the repository menu, the user can select cases associated with specific files. The set could then be linked to exploration page to view the variants associated with this set of cases.
- Users can now upload a custom gene list to the exploration page and leverage the GDC search and visualization features for cases and variants associated with the gene set.
- Filters added for the gene entity page. For example:
 - Clicking on a mutated gene from the project page will display mutations associated with the gene that are present in this project (filtered protein viewer, etc.).
 - Clicking on a mutated gene from the exploration page will display the mutations associated with the gene filtered by additional search criteria, such as "primary site is Kidney and mutation impact is high".
- UUIDs are now hidden from tables and charts to simplify readability. The UUIDs can still be exported and viewed in the tables using the "arrange columns" feature. In the mutation table, UUIDs are automatically exported.
- Mutation entity page one consequence per transcript is shown (10 rows by default) in the consequence table. The user should display all rows before exporting the table.

Bugs Fixed Since Last Release

- Exploration
 - Combining "Variant Caller" mutation filter with a case filter will display incorrect counts in the mutation facet. The number of mutations in the resulting mutation table is correct.
 - Mutation table: it is difficult to click on the denominator in "#Affected Cases in Cohort" column displayed to the left side of the bar. The user should click at a specific position at the top of the number to be able to go to the corresponding link.

Known Issues and Workarounds

- Visualizations
 - Data Portal graphs cannot be exported as PNG images in Internet Explorer. Graphs can be exported in PNG or SVG format from Chrome or Firefox browsers. Internet Explorer does not display chart legend and title when re-opening previously downloaded SVG files, the recommendation is to open downloaded SVG files with another program.
 - In the protein viewer there may be overlapping mutations. In this case mousing over a point will just show a single mutation and the other mutations at this location will not be apparent.
- Project page
 - On the project page, the Summary Case Count link should open the case tab on the Repository page instead it opens
 the file page
- Entity page
 - On the mutation entity page, in the Consequences Table, the "Coding DNA Change" column is not populated for rows
 that do not correspond to the canonical mutation.
- Repository and Cart
 - The annotation count in File table of Repository and Cart does not link to the Annotations page anymore. The user can navigate to the annotations through the annotation count in Repository Case table.

• Legacy Archive

- Downloading a token in the GDC Legacy Archive does not refresh it. If a user downloads a token in the GDC Data Portal and then attempts to download a token in the GDC Legacy Archive, an old token may be provided. Reloading the Legacy Archive view will allow the user to download the updated token.
- Exporting the Cart table in JSON will export the GDC Archive file table instead of exporting the files in the Cart only.

• Web Browsers

- Browsers limit the number of concurrent downloads, it is generally recommended to add files to the cart and download large number of files through the GDC Data Transfer Tool, more details can be found on GDC Website.
- The GDC Portals are not compatible with Internet Explorer running in compatibility mode. Workaround is to disable compatibility mode.

Release details are maintained in the GDC Data Portal Change Log.

Release 1.6.0

GDC Product: GDC Data Portal
Release Date: June 29, 2017

New Features and Changes

There was a major new release of the GDC Data Portal focused on Data Analysis, Visualization, and Exploration (DAVE). Some important new features include the following:

- New visual for the Homepage: a human body provides the number of Cases per Primary Site with a link to an advanced Cancer Projects search
- The Projects menu provides the Top 20 Cancer Genes across the GDC Projects and the Case Distribution per Project
- A new menu "Exploration" is an advanced Cancer Projects search which provides the ability to apply Case, Gene, and Mutation filters to look for:
 - List of Cases with the largest number of Somatic Mutations
 - The most frequently mutated Genes
 - The most frequent Variants
 - Oncogrid view of mutation frequency
- Visualizations are provided across the Project, Case, Gene and Mutation entity pages:
 - List of most frequently mutated genes and most frequent variants
 - Survival plots for patients with or without specific variants
 - Survival plots for patients with or without variants in specific genes
 - Lollipop plots of mutation frequency across protein domains
- Links to external databases (COSMIC, dbSNP, Uniprot, Ensembl, OMIM, HGNC)
- Quick Search for Gene and Mutation entity pages
- The ability to export the current view of a table in TSV
- Retired GDC cBioPortal

For detailed updates please review the Data Portal User Guide.

Bugs Fixed Since Last Release

• BAM Slicing dialog box does not disappear automatically upon executing the BAM slicing function. The box can be closed manually.

- Very long URLs will produce a 400 error. Users may encounter this after clicking on "source files" on a file page where the target file is derived from hundreds of other files such as for MAF files.
- If bam slicing produces an error pop-up message it will be obscured behind the original dialog box.
 - Internet Explorer users are not able to use the "Only show fields with no values" when adding custom facets
 - Exporting large tables in the Data Portal may produce a 500 error. Filtering this list to include fewer cases or files should eliminate the error

• New Visualizations

- Cannot export Data Portal graphs in PNG in Internet Explorer. Graphs can be exported to PNG or SVG from Chrome or Firefox browsers. Internet would not display chart legend and title when re-opening previously downloaded SVG files, recommendation is to open downloaded SVG files with another software.
- In the protein viewer there may be overlapping mutations. In this case mousing over a point will just show a single mutation and the other mutations at this location will not be apparent.

• Exploration

- Combining "Variant Caller" mutation filter with a case filter will display wrong counts in the mutation facet. The number of mutations in the result mutation table is correct.
- Mutation table: it is difficult to click on the denominator in "#Affected Cases in Cohort" column displayed to the left side of the bar. The user should click at a specific position at the top of the number to be able to go to the corresponding link.

• Entity page

- On the mutation entity page, in the Consequences Table, the "Coding DNA Change" column is not populated for rows that do not correspond to the canonical mutation.

• Repository and Cart

- The annotation count in File table of Repository and Cart does not link to the Annotations page anymore. The user can navigate to the annotations through the annotation count in Repository - Case table.

• Legacy Archive

- Downloading a token in the GDC Legacy Archive does not refresh it. If a user downloads a token in the GDC Data Portal and then attempts to download a token in the GDC Legacy Archive, an old token may be provided. Reloading the Legacy Archive view will allow the user to download the updated token.
- Exporting the Cart table in JSON will export the GDC Archive file table instead of exporting the files in the Cart only.

• Web Browsers

- Browsers limit the number of concurrent downloads, it is generally recommended to add files to the cart and download large number of files through the GDC Data Transfer Tool, more details can be found on GDC Website.
- The GDC Portals are not compatible with Internet Explorer running in compatibility mode. Workaround is to disable compatibility mode.

Release details are maintained in the GDC Data Portal Change Log.

Release 1.5.2

GDC Product: GDC Data Portal
Release Date: May 9, 2017

New Features and Changes

- Removed link to Data Download Statistics Report
- Updated version numbers of API, GDC Data Portal, and Data Release

• None

Known Issues and Workarounds

• General

- Exporting large tables in the Data Portal may produce a 500 error. Filtering this list to include fewer cases or files should eliminate the error
- After successful authentication, the authentication popup does not close for Internet Explorer users running in "Compatibility View". Workaround is to uncheck "Display Intranet sites in Compatibility View" in Internet Explorer options. Alternatively, refreshing the portal will correctly display authentication status.
- BAM Slicing dialog box does not disappear automatically upon executing the BAM slicing function. The box can be closed manually.
- Due to preceding issue, If bam slicing produces an error pop-up message it will be obscured behind the original dialog box.
- Very long URLs will produce a 400 error. Users may encounter this after clicking on "source files" on a file page where the target file is derived from hundreds of other files such as for MAF files. To produce a list of source files an API call can be used with the search parameter "fields=analysis.input_files.file_name".
 - * Downloading a token in the GDC Legacy Archive does not refresh it. If a user downloads a token in the GDC Data Portal and then attempts to download a token in the GDC Legacy Archive, an old token may be provided. Reloading the Legacy Archive view will allow the user to download the updated token.

Example

1 https://api.gdc.cancer.gov/files/455e26f7-03f2-46f7-9e7a-9c51ac322461?pretty=true&fields=analysis.input_files.fi

• Cart

 Counts displayed in the top right of the screen, next to the Cart icon, may become inconsistent if files are removed from the server.

• Web Browsers

- Browsers limit the number of concurrent downloads, it is generally recommended to add files to the cart and download large number of files through the GDC Data Transfer Tool, more details can be found on GDC Website.
- Internet Explorer users are not able to use the "Only show fields with no values" when adding custom facets
- The GDC Portals are not compatible with Internet Explorer running in compatibility mode. Workaround is to disable compatibility mode.

Release details are maintained in the GDC Data Portal Change Log.

Release 1.4.1

GDC Product: GDC Data Portal
Release Date: October 31, 2016

New Features and Changes

- Added a search feature to help users select values of interest in certain facets that have many values.
- Added support for annotation ID queries in quick search.
- Added a warning when a value greater than 90 is entered in the "Age at Diagnosis" facet.
- Added Sample Type column to file entity page.
- Authentication tokens are refreshed every time they are downloaded from the GDC Data Portal.

- Buttons are inactive when an action is in progress.
- Improved navigation features in the overview chart on portal homepage.
- Removed State/Status from File and Case entity pages
- Removed the "My Projects" feature.
- Removed "Created" and "Updated" dates from clinical and biospecimen entities.

- Advanced search did not accept negative values for integer fields.
- Moving from facet search to advanced search resulted in an incorrect advanced search query.
- Some facets were cut off in Internet Explorer and Firefox.

Known Issues and Workarounds

- General
 - Exporting large tables in the Data Portal may produce a 500 error. Filtering this list to include fewer cases or files should eliminate the error
 - After successful authentication, the authentication popup does not close for Internet Explorer users running in "Compatibility View". Workaround is to uncheck "Display Intranet sites in Compatibility View" in Internet Explorer options. Alternatively, refreshing the portal will correctly display authentication status.
 - BAM Slicing dialog box does not disappear automatically upon executing the BAM slicing function. The box can be closed manually.
 - Due to preceding issue, If bam slicing produces an error pop-up message it will be obscured behind the original dialog box.
 - Very long URLs will produce a 400 error. Users may encounter this after clicking on "source files" on a file page where the target file is derived from hundreds of other files such as for MAF files. To produce a list of source files an API call can be used with the search parameter "fields=analysis.input_files.file_name".
 - * Downloading a token in the GDC Legacy Archive does not refresh it. If a user downloads a token in the GDC Data Portal and then attempts to download a token in the GDC Legacy Archive, an old token may be provided. Reloading the Legacy Archive view will allow the user to download the updated token.

Example

1 https://api.gdc.cancer.gov/files/455e26f7-03f2-46f7-9e7a-9c51ac322461?pretty=true&fields=analysis.input_files.fi

- Cart
 - Counts displayed in the top right of the screen, next to the Cart icon, may become inconsistent if files are removed from the server.
- Web Browsers
 - Browsers limit the number of concurrent downloads, it is generally recommended to add files to the cart and download large number of files through the GDC Data Transfer Tool, more details can be found on GDC Website.
 - Internet Explorer users are not able to use the "Only show fields with no values" when adding custom facets
 - The GDC Portals are not compatible with Internet Explorer running in compatibility mode. Workaround is to disable compatibility mode.

Release details are maintained in the GDC Data Portal Change Log.

Release 1.3.0

GDC Product: GDC Data Portal
Release Date: September 7, 2016

New Features and Changes

- A new "Metadata" button on the cart page to download merged clinical, biospecimen, and file metadata in a single consolidated JSON file. May require clearing browser cache
- Added a banner on the Data Portal to help users find data
- Added support for "Enter" key on login button
- On the Data page, the browser will remember which facet tab was selected when hitting the "Back" button
- In file entity page, if there is a link to one single file, redirect to this file's entity page instead of a list page.

Bugs Fixed Since Last Release

- Adding a mix of open and controlled files to the cart from any Case entity pages was creating authorization issues
- Opening multiple browser tabs and adding files in those browser tabs was not refreshing the cart in other tabs.
- When user logs in from the advanced search page, the login popup does not automatically close
- When removing a file from the cart and clicking undo, GDC loses track of permission status of the user towards this file and will ask for the user to log-in again.
- Download File Metadata button produces incomplete JSON output omitting such fields as file_name and submitter_id. The current workaround includes using the API to return file metadata.
- Annotations notes do not wrap to the next line at the beginning or the end of a word, some words might be split in two lines
- Sorting annotations by Case UUID causes error

Known Issues and Workarounds

- General
 - When no filters are engaged in the Legacy Archive or Data Portal, clicking the Download Manifest button may produce a 500 error and the message "We are currently experiencing issues. Please try again later.". To avoid this error the user can first filter by files or cases to reduce the number files added to the manifest.
 - After successful authentication, the authentication popup does not close for Internet Explorer users running in "Compatibility View". Workaround is to uncheck "Display Intranet sites in Compatibility View" in Internet Explorer options. Alternatively, refreshing the portal will correctly display authentication status.
 - BAM Slicing dialog box does not disappear automatically upon executing the BAM slicing function. The box can be closed manually.
 - Due to preceding issue, If bam slicing produces an error pop-up message it will be obscured behind the original dialog box.
 - Very long URLs will produce a 400 error. Users may encounter this after clicking on "source files" on a file page where the target file is derived from hundreds of other files such as for MAF files. To produce a list of source files an API call can be used with the search parameter "fields=analysis.input_files.file_name".
 - On the Legacy Archive, searches for "Case Submitter ID Prefix" containing special characters are not displayed correctly above the result list. The result list is correct, however.

Example

$1\ \texttt{https://api.gdc.cancer.gov/files/455e26f7-03f2-46f7-9e7a-9c51ac322461?} \\ \texttt{pretty=true\&fields=analysis.input_files.fil$

- Cart
 - Counts displayed in the top right of the screen, next to the Cart icon, may become inconsistent if files are removed from the server.
- Web Browsers
 - Browsers limit the number of concurrent downloads, it is generally recommended to add files to the cart and download large number of files through the GDC Data Transfer Tool, more details can be found on GDC Website.
 - Internet Explorer users are not able to use the "Only show fields with no values" when adding custom facets
 - The GDC Portals are not compatible with Internet Explorer running in compatibility mode. Workaround is to disable compatibility mode.

Release 1.2.0

GDC Product: GDC Data Portal
Release Date: August 9th, 2016

New Features and Changes

- Added a retry (1x) mechanism for API calls
- Added support for ID fields in custom facets
- Added Case Submitter ID to the Annotation entity page
- Added a link to Biospeciment in the Case entity page

Bugs Fixed Since Last Release

- General.
 - Not possible to use the browser's back button after hitting a 404 page
 - 404 page missing from Legacy Archive Portal
 - Table widget icon and export JSON icon should be different
 - Download SRA XML files from the legacy archive portal might not be possible in some context
- Data and facets
 - Default values for age at diagnosis is showing 0 to 89 instead of 0 to 90
 - Biospecimen search in the case entity page does not highlight (but does bold and filter) results in yellow when title
 case is not followed
 - Table sorting icon does not include numbers
 - '-' symbol is missing on empty fields (blank instead), additional missing fields identified since last release. ### Known Issues and Workarounds

• General

- When no filters are engaged in the Legacy Archive or Data Portal, clicking the Download Manifest button may produce a 500 error and the message "We are currently experiencing issues. Please try again later.". To avoid this error the user can first filter by files or cases to reduce the number files added to the manifest.
- After successful authentication, the authentication popup does not close for Internet Explorer users running in "Compatibility View". This only impact users at the NIH. Workaround is to uncheck "Display Intranet sites in Compatibility View" in Internet Explorer options. Alternatively, refreshing the portal will correctly display authentication status.
- When user login from the advanced search page, the login popup does not automatically close

• Cart

- When removing a file from the cart and clicking undo, GDC looses track of permission status of the user towards this file and will ask for the user to log-in again.
- Counts displayed in the top right of the screen, next to the Cart icon, might get inconsistent if files are removed from the server.
- Download File Metadata button produces incomplete JSON output omitting such fields as file_name and submitter_id.
 The current workaround includes using the API to return file metadata.

• Annotations

- Annotations notes do not wrap to the next line at the beginning or the end of a word, some words might be split in two lines
- Sorting annotations by Case UUID causes error

• Web Browsers

- Browsers limit the number of concurrent downloads, it is generally recommended to add files to the cart and download large number of files through the GDC Data Transfer Tool, more details can be found on GDC Website.
- Internet Explorer users are not able to use the "Only show fields with no values" when adding custom facets

 The GDC Portals are not compatible with Internet Explorer running in compatibility mode. Workaround is to disable compatibility mode

Release details are maintained in the GDC Data Portal Change Log.

Release 1.1.0

GDC Product: GDC Data Portal
Release Date: June 1st, 2016

New Features and Changes

• This is a bug-fixing release, no new features were added.

Bugs Fixed Since Last Release

- General
 - Fixed 508 compliance issues.
 - Disabled download manifest action on projects without files.
 - Updated the portal to indicate to the user that his session expired when he tries to download the authentication token.
 - Unselected "My project" filter after user logs-in.
 - Fixed missing padding when query includes "My Projects".
 - Enforced "Add to cart" limitation to 10,000 files everywhere on the Data Portal.
- Tables
 - Improved usability of the "Sort" feature
 - Updated the "Add all files to cart" button to add all files corresponding to the current query (and not only displayed files).
 - Fixed an issue where Platform would show "0" when selected platform is "Affymetrix SNP 6.0".
- Data
 - Corrected default values populated when adding a custom range facet.
 - Fixed an issue preventing the user to sort by File Submitter ID in data tables.
- File Entity Page
 - Improved "Associated Cases/Biospecimen" table for files associated to a lot of cases.
 - Fixed an error when performing BAM Slicing.

Known Issues and Workarounds

- General.
 - After successful authentication, the authentication popup does not close for Internet Explorer users running in "Compatibility View". This only impact users at the NIH. Workaround is to uncheck "Display Intranet sites in Compatibility View" in Internet Explorer options. Alternatively, refreshing the portal will correctly display authentication status.
 - Download SRA XML files from the legacy archive portal might not be possible in some context
 - Not possible to use the browser's back button after hitting a 404 page
 - 404 page missing from Legacy Archive Portal
 - Table widget icon and export JSON icon should be different
- Data and facets
 - Default values for age at diagnosis is showing 0 to 89 instead of 0 to 90

- Biospecimen search in the case entity page does not highlight (but does bold and filter) results in yellow when title
 case is not followed
- Table sorting icon does not include numbers
- '-' symbol is missing on empty fields (blank instead), additional missing fields identified since last release.

• Cart

- When removing a file from the cart and clicking undo, GDC looses track of permission status of the user towards this file and will ask for the user to log-in again.
- Counts displayed in the top right of the screen, next to the Cart icon, might get inconsistent if files are removed from the server.

• Annotations

Annotations notes do not wrap to the next line at the beginning or the end of a word, some words might be split in
two lines

• Web Browsers

- Browsers limit the number of concurrent downloads, it is generally recommended to add files to the cart and download large number of files through the GDC Data Transfer Tool, more details can be found on GDC Website.
- Internet Explorer users are not able to use the "Only show fields with no values" when adding custom facets
- The GDC Portals are not compatible with Internet Explorer running in compatibility mode. Workaround is to disable compatibility mode

Release details are maintained in the GDC Data Portal Change Log.

Release 1.0.1

GDC Product: GDC Data Portal
Release Date: May 18, 2016

New Features and Changes

• This is a bug-fixing release, no new features were added.

Bugs Fixed Since Last Release

- Tables and Export
 - Restore default table column arrangement does not restore to the default but it restores to the previous state
- Cart and Download
 - Make the cart limit warning message more explanatory
 - In some situations, adding filtered files to the cart might fail
- Layout, Browser specific and Accessibility
 - When disabling CSS, footer elements are displayed out of order
 - If javascript is disabled html tags are displayed in the warning message
 - Layout issues when using the browser zoom in function on tables
 - Cart download spinner not showing at the proper place
 - Not all facets are expanded by default when loading the app

• General

- If a user has previously logged into the Portal and left a session without logging out, if the user returns to the Portal after the user's sessionID expires, it looks as if the user is still authenticated. The user cannot download the token and gets an error message that would not close. The user should clear the cache to properly log out.
- '-' symbol is missing on empty fields (blank instead)
- Download manifest button is available for TARGET projects with 0 files, resulting in error if user clic on button
- After successful authentication, the authentication popup does not close for Internet Explorer users running in "Compatibility View". This only impact users at the NIH. Workaround is to uncheck "Display Intranet sites in Compatibility View" in Internet Explorer options. Alternatively, refreshing the portal will correctly display authentication status.

• Data

- When adding a custom range facet, default values are incorrectly populated
- The portal might return incorrect match between cases and files when using field cases.samples.portions.created_datetime (custom facet or advanced search). Note: this is not a UI issue.
- Sorting File Submitter ID option on the file tab result in a Data Portal Error
- Tables and Export
 - Table sorting icon does not include numbers
- Browsers limit the number of concurrent downloads, it is generally recommended to add files to the cart and download large number of files through the GDC Data Transfer Tool, more details can be found on GDC Website.